

REMARKS

Included herewith is a Request for Continued Examination (“RCE”) along with a three month petition of time to respond to the Office Action dated October 3, 2007. Deposit Account 20-0823 may be charged a fee of \$405 (fee code 2801) for the request for continued examination and \$525 (fee code 2253) for the extension fee. Also, in the event that any additional fees are necessary, such fees are hereby authorized to be charged to our Deposit Account 20-0823.

In this response, Applicant has amended Claims 23, 25, 26, 27-31, 35, 37-39, 41-44, and 52-55. Claims 1-22, 24, 32, 34, 45, 47, and 48-51 were previously canceled, and Claims 33, 36, 40 and 46 are currently canceled. New Claims 56, 57 and 58 are system claim equivalents to Claims 27, 28 and 29. New Claim 59 is a method claim equivalent to Claim 44. These new claims, i.e., Claims 56-59 are added for symmetry and are respectfully believed to overcome all of rejections noted above in the same manner as Claims 27, 28, 29 and 44. No new matter has been added. Thus, Claims 23, 25, 26, 27-31, 35, 37-39, 41-44, and 52-59 are pending in this patent application for the Examiner’s consideration.

Rejections under 35 U.S.C. § 112:

Claim 23 was rejected under 35 U.S.C. § 112 for failing to comply with the written description requirement and for failing to set forth that which the inventor regards as the invention. Claim 23 is amended to recite: “A method of utilizing an adaptive speaker identity verification system comprising:....” Support for this amendment can be found on Page 2, Paragraph [0016], Lines 1-2 of Applicant’s Published Patent Application, i.e., U.S. Patent Application Publication No. 20020198857, published December 26, 2002, which recites: “As a second example, consider an **adaptive speaker identity verification system....**” (emphasis added). Therefore, no new matter has been added. An adaptive speaker identity verification

system is a well-known commercially available device. For example, a speaker identity verification system is described in detailed in U.S. Patent No. 5,517,558, issued to Schalk, see Exhibit A, which recites: “This method is implemented according to the invention using a system comprising a digital processor, storage means connected to the digital processor, prompt means controlled by the digital processor for prompting a caller to speak a password beginning with a first digit and ending with a last digit thereof, speech processing means controlled by the digital processor for effecting a multistage data reduction process and generating resultant voice recognition and voice verification parameter data, and voice recognition and verification decision routines.” (Schalk, Column 2, Lines 9-18). Also enclosed as Exhibit B is “*Automatic Speaker Recognition Recent Progress, Current Applications, and Future Trends*”, which was Presented at the AAAS 2000 Meeting, Humans, Computers and Speech Symposium on February 19, 2000, shows this term is for a well-known computerized device that is a programmable, physical machine with a defined meaning. There are numerous technical papers on this type of device, e.g., Rosenberg, A.; Sambur, M., “New techniques for automatic speaker verification,” *Acoustics, Speech, and Signal Processing [see also IEEE Transactions on Signal Processing]*, Vol. 23, No. 2, pp. 169-176, April 1975, see Summary in Exhibit C as attached. Moreover, this term is so well known in the art for a physical machine, having a computer, that it is now a recognized acronym, i.e., ASV, for automatic speaker verification, See Exhibit D.

Amended Claim 23 also now recites: “...receiving first input data, which represents a person’s unclassified speech utilizing the adaptive speaker identity verification system;...” Support for this amendment can be found on Page 3, Paragraph [0039], Line 1 of Applicant’s Published Patent Application, i.e., U.S. Patent Application Publication No. 20020198857,

published December 26, 2002, which recites: “In operation on **unclassified** input items 67,...” (emphasis added). Therefore, no new matter has been added.

Moreover, amended Claim 23 now recites: “...receiving second input data, which represents in part probability distributions for authentic and spurious classes based upon the pooled output statistics of the adaptive speaker identity verification system, including the equal error rate, and which represents in part optional parameters to focus on at least one region of interest in a decision space;...” Support for this amendment can be found on Page 2, Paragraph [0035], Lines 1-11, Page 3, Paragraph [0037], Lines 1-7, and Page 3, Paragraph [0049], Lines 1-14 of Applicant’s Published Patent Application, i.e., U.S. Patent Application Publication No. 20020198857, published December 26, 2002, which recites: “Normalized Detector Scaling (NDS) represents a means of providing context independent decision rules 50 for operating a pattern recognition system 51. NDS also provides the user of a pattern recognition system a simpler means of controlling the decision criterion. This comes at the cost of an additional complexity in the pattern recognition system 51, as compared to either parametric 11, or non-parametric 21 pattern recognition systems. **The pattern recognition system must be able to provide output statistics 61 for the authentic 31 and spurious 32 class-specific probability distributions.**” “The NDS transform constructor 62 takes as input the **pooled output statistics 61**, or the **probability distributions** of the pattern recognition system. The NDS transform constructor 62 also takes as input optional transform parameters 65 that may serve, for example, to tailor or focus the NDS transform on a **particular region of interest in the decision space.**” “Other methods for combining information from both the authentic and spurious probability distributions are possible. One such method produces a scale with two regions. **The regions are formed by the EER criterion**, and represent the likelihood of a test item belonging to a

particular class. The first region refers to test items unlikely to be authentic, and is simply a mapping onto a scale linear in probability, as described above, of the cumulative probability distribution from $-\infty$ to the EER criterion of the spurious class output statistics. The second region refers to test items likely to be authentic, and is simply a mapping onto a scale linear in probability, as described above, of the cumulative probability distribution from the EER criterion to ∞ of the authentic class output statistics.” (emphasis added). Therefore, no new matter has been added.

In addition, amended Claim 23 also now recites: “...computing a transform of the first input data using the second input data with a normalized detector scale transformer associated with the adaptive speaker identity verification system onto a normalized, one dimension, decision scale based on the transform; and....” Support for this amendment can be found on Page 2, Paragraph [0020], Lines 1-8 and Page 3, Paragraph [0039], Lines 1-4 of Applicant’s Published Patent Application, i.e., U.S. Patent Application Publication No. 20020198857, published December 26, 2002, which recites: “More specifically, an object of the present invention is to provide a **Normalized Detector Scaling** method that utilizes the class-specific probability distributions of a pattern recognition system to make the selection of the operating criteria independent of the particulars of the pattern recognition system. This being accomplished by **transforming** the pattern recognition system output statistics to a well-defined, **one-dimensional scale**.” “In operation on unclassified input items 67, the pattern recognition system output statistics 66 are presented to the **NDS transformer 64** that uses the NDS transform 63 to convert the output statistics 66 to the new decision space.” (emphasis added). Therefore, no new matter has been added.

Finally, amended Claim 23 also now recites: "...establishing at least one decision criterion, wherein the at least one decision criterion corresponds to a level of similarity or a level of dissimilarity between the first input data representing a person's unclassified speech data and the second input data with the adaptive speaker identity verification system." Support for this amendment can be found on Page 3, Paragraph [0040], Lines 1-11 and Paragraph [0041], Lines 1-5 of Applicant's Published Patent Application, i.e., U.S. Patent Application Publication No. 20020198857, published December 26, 2002, which recites: "The NDS transform constructor 62 relies on the pattern recognition system's pooled output statistics 61, which are essentially represented by the probability distributions for the authentic 31 and spurious 32 classes. If these output statistics 61 represent **dissimilarities**, i.e. numbers that increase as the match to a known class decreases, the dissimilarities d , are converted to **similarities** s , so that the intuitive notion of "bigger is better" is utilized. This can be done as simply as $s = d_{\max} - d$. FIG. 7 illustrates the authentic 71 and spurious 72 distributions of FIG. 3 converted from a scale of dissimilarity to a scale of **similarity**." "Information from both the authentic 71 and spurious 72 probability distributions are combined by some method to sufficiently simplify the **decision criteria selection** so that only a single number has to be selected for operation of the pattern recognition system." (emphasis added). Therefore, no new matter has been added.

It is respectfully believed that the amended language of Claim 23 is fully and completely disclosed in Applicant's Published Patent Application, i.e., U.S. Patent Publication No. 20020198857, published December 26, 2002.

Therefore, it is respectfully believed that the rejection of Claim 23 under 35 U.S.C. § 112 is overcome.

Claims 25-31, 33, 52 and 53 were also rejected under 35 U.S.C. § 112 for failing to comply with the written description requirement and for failing to set forth that which the inventor regards as the invention since these Claims depend from Claim 23 (Office Action, Page 4, Lines 1-3). Claim 33 is now canceled and it is respectfully believed that this rejection is rendered moot with respect to Claim 33. Since Claims 25-31, 52 and 53 depend from Claim 23 and contain all of the limitations of Claim 23, as amended, it is respectfully believed that Claims 25-31, 52 and 53 overcome the rejection under 35 U.S.C. § 112 in the same manner as Claim 23.

Claim 35 was rejected under 35 U.S.C. § 112 for failing to comply with the written description requirement and for failing to set forth that which the inventor regards as the invention. Claim 35, as now amended, is a system claim version of method Claim 23 and overcomes the rejection under 35 U.S.C. § 112 in the same manner as Claim 23. Claims 36-44, 46, 54 and 55 were also rejected under 35 U.S.C. § 112 for failing to comply with the written description requirement and for failing to set forth that which the inventor regards as the invention since these Claims depend from Claim 35 (Office Action, Page 4, Lines 4-14). Claims 36, 40 and 46 are now canceled and it is respectfully believed that this rejection is rendered moot with respect to Claims 36, 40 and 46. Since Claims 37-39, 41-44, 54 and 55 depend from Claim 35 and contain all of the limitations of Claim 35, as amended, it is respectfully believed that Claims 37-39, 41-44, 54 and 55 overcome the rejection under 35 U.S.C. § 112 in the same manner as Claim 35.

Claims 23, 25-31, 33, 52 and 53 and Claims 35-44, 46, 54 and 55 are rejected under 35 U.S.C. § 112 for failing to comply with the enablement requirement. Claims 33, 36, 40 and 46 are now canceled and it is respectfully believed that this rejection is rendered moot with respect to Claims 33, 36, 40 and 46. Declarations of David P. Morgan, who is the Vice President,

Enterprise Technology & Architecture of Fidelity Investments Systems Company in Boston, Massachusetts, is attached as Exhibit E and of Michael Phillips, who is the Co-Founder and Chief Technology Officer of Vlingo Corp., in Cambridge, Massachusetts, and is the Co-Founder of SpeechWorks International, Inc. in 1994 (currently Nuance Communications, Inc., Boston, Massachusetts), is attached as Exhibit F. Michael Phillips has been active in the speech technology world for over twenty years. Michael started his career as a researcher first at Carnegie Mellon University and then at the Spoken Language Systems group at Massachusetts Institute of Technology (“MIT”) working on core technology for automatic speech recognition. In 1994, Michael founded SpeechWorks International, Inc. based on technology that Michael and others had developed at MIT. Over the next ten years, Michael and team grew SpeechWorks International, Inc. from a small startup in a new market into the market leader in the now established market for speech enabled call center solutions. SpeechWorks International, Inc. was responsible for many of key innovations in use today in the speech recognition systems deployed throughout the world. In 2003, SpeechWorks International, Inc. was acquired by ScanSoft Inc. (now named Nuance Communications, Inc.). Michael joined ScanSoft Inc. as CTO and oversaw technology integration and development across the product groups. In 2005, Michael left ScanSoft Inc. to spend a year as a visiting scientist at MIT before starting Vlingo Corp. in the summer of 2006.

Both Experts believe that an “adaptive speaker identity verification system” is well known in the art for a physical machine, having a computer, which receives a person’s unclassified speech and converts that speech to data and then is able to perform analysis on that data utilizing statistics to verify the identity of a particular person. Moreover, both Experts believe that a person skilled in speaker identity verification technology would easily be

able to implement the Applicant's Invention disclosed in U.S. Patent Application Publication No. 20020198857 in an adaptive speaker identity verification system by merely reading U.S. Patent Application Publication No. 20020198857 and then programming the adaptive speaker identity verification system. Both individuals believe that it would be **a very straightforward process** based on a reading of U.S. Patent Application Publication No. 20020198857, so there would be no need for any undue experimentation involving the adaptive speaker identity verification system. Therefore, there are two Declarations from Experts in this field that believe that it would be a very simple and straightforward process to program a commonly available adaptive speaker identity verification system to replicate the features found in Applicant's U.S. Patent Application Publication No. 20020198857, which includes the limitations found in Claims 23, 25-31, 35, 37-39, 41-44, and 52-55. Therefore, it is respectfully believed that Claims 23, 25-31, 35, 37-39, 41-44, and 52-55 overcome the rejection under 35 U.S.C. § 112 by being fully and completely enabled.

Rejections under 35 U.S.C. § 101:

Claims 23, 25-31, 33, 35-44, 46, and 52-55 were rejected under 35 U.S.C. § 101 for reciting a mathematical algorithm. Claims 33, 36, 40 and 46 are canceled and it is respectfully believed that this rejection with regard to Claims 33, 36, 40 and 46 is rendered moot. Claims 23 and 35 are amended to recite an **"adaptive speaker identity verification system"** (emphasis added). Support for this amendment can be found on Page 2, Paragraph [0016], Lines 1-2 of Applicant's Published Patent Application, i.e., U.S. Patent Publication No. 20020198857, published December 26, 2002, which recites: "As a second example, consider an **adaptive speaker identity verification system....**" (emphasis added). Therefore no new matter is added.

An adaptive speaker identity verification system is a well-known commercially available device. For example, a speaker identity verification system is described in detailed in U.S. Patent No. 5,517,558, issued to Schalk, see Exhibit A, which recites: “This method is implemented according to the invention using a system comprising a digital processor, storage means connected to the digital processor, prompt means controlled by the digital processor for prompting a caller to speak a password beginning with a first digit and ending with a last digit thereof, speech processing means controlled by the digital processor for effecting a multistage data reduction process and generating resultant voice recognition and voice verification parameter data, and voice recognition and verification decision routines.” (Schalk, Column 2, Lines 9-18). Also enclosed as Exhibit B is “*Automatic Speaker Recognition Recent Progress, Current Applications, and Future Trends*,” which was Presented at the AAAS 2000 Meeting, Humans, Computers and Speech Symposium on February 19, 2000, shows this term is for a well-known computerized device that is a programmable, physical machine with a defined meaning. There are numerous technical papers on this type of device, e.g., Rosenberg, A.; Sambur, M., “New techniques for automatic speaker verification,” *Acoustics, Speech, and Signal Processing [see also IEEE Transactions on Signal Processing]*, Vol. 23, No. 2, pp. 169-176, April 1975, see Summary in Exhibit C as attached. Moreover, this term is so well known in the art for a physical machine, having a computer, that it is now a recognized acronym, i.e., ASV, for automatic speaker verification, See Exhibit D.

Also, as previously stated, there are Declarations of David P. Morgan, who is the Vice President, Enterprise Technology & Architecture of Fidelity Investments Systems Company in Boston, Massachusetts, which is attached as Exhibit E, and of Michael Phillips, who is the Co-Founder and Chief Technology Officer of Vlingo Corp., in Cambridge, Massachusetts, and is the

Co-Founder of SpeechWorks International, Inc., which is attached as Exhibit F. Both individuals believe that an “adaptive speaker identity verification system,” **is well-known in the art for a physical machine, having a computer, which receives a person’s unclassified speech and converts that speech to data and then is able to perform analysis on that data utilizing statistics to verify the identity of a particular person.** Moreover, both Experts believe that a person skilled in speaker identity verification technology would easily be able to implement the Applicant’s Invention disclosed in U.S. Patent Application Publication No. 20020198857 in an adaptive speaker identity verification system by merely reading U.S. Patent Application Publication No. 20020198857 and then programming the adaptive speaker identity verification system. Both individuals believe that it would be **a very straightforward process** based on a reading of U.S. Patent Application Publication No. 2002/0198857, so there would be no need for any undue experimentation involving the adaptive speaker identity verification system. Therefore, there are two Declarations from Experts in this field who believe that it would be a very simple and straightforward process to program a commonly available adaptive speaker identity verification system to replicate the features found in Applicant’s U.S. Patent Application Publication No. 20020198857, which includes the limitations found in Claims 23 and 35.

Since Claims 23 and 35 provide limitations of a **physical machine, having a computer,** directly in each Claim, it is respectfully believed this rejection under 35 U.S.C. § 101 is overcome. Under the Manual for Patent Examining Procedure (“M.P.E.P”) § 2107, the Examiner is required to “...ensure that the claims define statutory subject matter i.e., a process, **machine**, manufacture, composition of matter, or improvement thereof.” (emphasis added). In this case, **an adaptive speaker identity verification system is a well known machine that**

utilizes computer. Moreover, “if at any time during the examination, it becomes readily apparent that the claimed invention has a well-established utility, do not impose a rejection based on lack of utility. An invention has a well-established utility if (i) a person of ordinary skill in the art would immediately appreciate why the invention is useful based on the characteristics of the invention (e.g., properties or applications of a product or process), and (ii) the utility is specific, substantial, and credible.” M.P.E.P § 2107. In this case, an adaptive speaker identity verification system that is programmed to perform the features found in Claims 23 and 35 provides a very **specific, substantial, and credible utility** by providing a simple, intuitive, one-dimensional, decision support scale that is completely independent of the underlying features of the adaptive speaker verification system to assist in controlling the decision criterion in a wide variety of verbal transactions, e.g., financial transactions.

Therefore, since an “adaptive speaker identity verification system” is a well-known physical machine that utilizes a computer, and this term was present in the Applicant’s original patent application, as filed, it is respectfully believed that “**it is clear within which of the enumerated categories a claimed invention falls,**” i.e., **machine**, and Claims 23 and 35 overcome the rejection under 35 U.S.C. § 101.

Since Claims 25-31, 37-39, 41-44, and 52-55 depend from Claims 23 and 35 and contain all of the limitations of Claims 23 and 35, as amended, it is respectfully believed that Claims 25-31, 37-39, 41-44, and 52-55 overcome the rejection under 35 U.S.C. § 101 in the same manner as Claims 23 and 35.

Rejections under 35 U.S.C. § 103(a):

Claim 23 and Claim 35 were rejected under Hamid (U.S. Patent No. 6,038,334) in view of Campbell (“*Object Recognition for an Intelligent Room*, IEEE Conference on Computer

Vision and Pattern Recognition, Hilton Head, South Carolina, June 2000”). Claim 23 has been amended to now recite “receiving first input data, which represents a person’s unclassified speech utilizing the adaptive speaker identity verification system; receiving second input data, which represents in part probability distributions for authentic and spurious classes based upon the pooled output statistics of the adaptive speaker identity verification system, including the equal error rate, and which represents in part optional parameters to focus on at least one region of interest in a decision space;...” As shown in detail by Applicant in response to the rejection under 35 U.S.C. § 112, no new matter has been added. In marked contrast, Hamid recites: “A method of registering biometric information of an individual comprising the steps of: a) providing a biometric information sample from each of a plurality of different biometric sources of the same individual to at least one biometric input device in communication with a host processor; b) associating each provided biometric information sample with a biometric source; c) using the processor, registering each biometric information sample **against a template associated with the associated biometric source;**...” (Claim 1, Column 13, Lines 43-53) (emphasis added). Comparing a single data source against a template is very different operation than two sources of data. Moreover, the portions recited by the Examiner, Column 10, Line 48 to Column 11, Line 39 are directed to equations involving fingerprints and not speech. Moreover, Claim 23 further recites: “...computing a transform based on the output; and of the first input data using the second input data with a normalized detector scale transformer associated with the adaptive speaker identity verification system onto a normalized, one dimension, decision scale based on the transform;...” As shown in detail by Applicant in response to the rejection under 35 U.S.C. § 112, no new matter has been added. Also, Claim 23 further recites: “establishing at least one decision criterion, wherein the at least one decision criterion corresponds to a level of

similarity or a level of dissimilarity between the first input data representing a person's unclassified speech data and the second input data with the adaptive speaker identity verification system." As shown in detail by Applicant in response to the rejection under 35 U.S.C. § 112, no new matter has been added. These two features are wholly absent from Hamid. Moreover, Hamid simply obtains the biometric information and determines a coordinate in n Dimensional space and then determines a region in n Dimension space and then ascertains if the coordinate falls within that region as shown in Fig 5. Therefore, Hamid simply represents an example of a "non-parametric pattern recognition system."

Campbell et al. is directed to "...a new object recognition algorithm that is especially suited for finding everyday objects in an intelligent environment monitored by color video cameras" (Campbell et al., Page 1, Column 1, Section 1, Lines 1-4). Moreover, Campbell et al. recites: "We present an algorithm that can be trained with only a few images of the object, that requires only two parameters to be set, and that runs at 0.7 Hz on a normal PC with a normal color camera. The algorithm represents an object's features as small, quantized edge templates, and it represents the object's geometry with "Hough kernels". The Hough kernels implement a variant of the generalized Hough transform **using simple, 2D image correlation**. The algorithm also uses color information to **eliminate parts of the image from consideration**." (Campbell et al., Abstract, Page 1, Column 1, Lines 8-18) (emphasis added). Therefore, Campbell et al. is for a visual recognition system that teaches away from the Applicant's Invention by using objects and creating a two dimension correlation while the Applicant's Invention, as claimed, requires, "...computing a transform of the **first input data** using the **second input data** with a normalized detector scale transformer associated with the adaptive speaker identity verification system onto **a normalized, one dimension, decision scale** based on the transform."(emphasis added).

Therefore, Campbell et al. clearly teaches away from the Applicant's Invention that utilizes "**a normalized, one dimension, decision scale**" as claimed. The Supreme Court held in *U.S. v. Adams*, 383 U.S. 39, 148 U.S.P.Q. 479 (1966), that one important indicium of nonobviousness is "teaching away" from the claimed invention by the prior art or by experts in the art at (and/or after) the time the invention was made. This is specifically mandated by the Manual of Patent Examining Procedure (M.P.E.P.) § 2141.02, which recites: "A prior art reference must be considered in its entirety, i.e., as a whole, including portions that would lead away from the claimed invention." *W.L. Gore & Associates, Inc. v. Garlock, Inc.*, 721 F.2d 1540, 220 U.S.P.Q. 303 (Fed. Cir. 1983), cert. denied, 469 U.S. 851 (1984). Moreover, "...if proposed modification would render the prior art invention being modified unsatisfactory for its intended purpose, then there is no suggestion or motivation to make the proposed modification." *In re Gordon*, 733 F.2d 900, 221 U.S.P.Q. 1125 (Fed. Cir. 1984). The mere fact that references can be combined or modified does not render the resultant combination obvious unless the prior art also suggests the desirability of the combination. *In re Mills*, 916 F.2d 680, 16 U.S.P.Q.2d 1430 (Fed. Cir. 1990).

Moreover, it is respectfully believed to be axiomatic that this feature is not disclosed in either Hamid or Campbell et al., i.e., computing a transform of the **first input data** using the **second input data** with a normalized detector scale transformer associated with the adaptive speaker identity verification system onto **a normalized, one dimension, decision scale** based on the transform, cannot come into being by their combination. Moreover, there is no teaching, suggestion, or motivation in the prior art that would have led one of ordinary skill to modify the prior art reference or to combine prior art reference teachings to arrive at the claimed invention. "To reject a claim based on this rationale, U.S. Patent Office personnel must resolve the Graham factual inquiries. Office personnel must then articulate the following: (1) **a finding that the**

prior art included each element claimed, although not necessarily in a single prior art reference, with the only difference between the claimed invention and the prior art being the lack of actual combination of the elements in a single prior art reference; (2) a finding that one of ordinary skill in the art could have combined the elements as claimed by known methods, and that in combination, **each element merely would have performed the same function as it did separately.**” (Federal Register / Volume 72, No. 195 / Wednesday, October 10, 2007 / Notices, Page 57529, “*Examination Guidelines for Determining Obviousness Under 35 U.S.C. § 103 in View of the Supreme Court Decision in KSR International Co. v. Teleflex Inc.*”) (emphasis added).

The Applicant’s Invention, as claimed, requires the use of output statistics with an adaptive speaker verification system to provide a simple, intuitive, one-dimensional, decision support scale that is completely independent of the underlying features of the adaptive speaker verification system. Therefore, Hamid and Campbell et al. both describe methods in which the output of their pattern recognition system is a Yes/No decision based upon comparing a specific instance of a newly classified biometric image (or an visual object in Campbell et al.) against a pre-determined criterion or pre-determined limits, which is described in the Background of the Invention for the Applicant’s Published Patent Application, i.e., U.S. Patent Application Publication No. 20020198857, which is summarized as “traditional methods such as operating characteristic analysis.” (U.S. Patent Application Publication No. 20020198857, Page 2, Paragraph [0019], Line 4). In marked contrast, the Applicant’s invention is a unique, two-stage process that “utilizes the class-specific probability distribution of a pattern recognition system to make the selection of the operating criteria independent of the particulars of the pattern recognition system”, and that provides an intuitive interface for decision criteria selection. In

determining the differences between the prior art and the claims, the question under 35 U.S.C. § 103 is not whether the differences themselves would have been obvious, but whether the claimed invention as a whole would have been obvious. *Stratoflex, Inc. v. Aeroquip Corp.*, 713 F.2d 1530, 218 U.S.P.Q. 871 (Fed. Cir. 1983); *Schenck v. Nortron Corp.*, 713 F.2d 782, 218 U.S.P.Q. 698 (Fed. Cir. 1983).

Moreover, any statement that modifications of the prior art to meet the claimed invention would have been well within the ordinary skill of the art at the time the claimed invention was made because the references relied upon teach that all aspects of the claimed invention were individually known in the art is not sufficient to establish a *prima facie* case of obviousness without some objective reason to combine the teachings of the references. *Ex parte Levengood*, 28 U.S.P.Q.2d 1300 (Bd. Pat. App. & Inter. 1993). “[R]ejections on obviousness cannot be sustained by mere conclusory statements; instead, there must be some articulated reasoning with some rational underpinning to support the legal conclusion of obviousness.” *KSR International Co. v. Teleflex Inc.*, 82 U.S.P.Q.2d 1385 at 1396 (U.S. 2007) quoting *In re Kahn*, 441 F.3d 977, 988, 78 U.S.P.Q.2d 1329, 1336 (Fed. Cir. 2006). There is no reason to modify the two dimensional visual object recognition system of Campbell or the fingerprint recognition system in n dimensional space of Hamid to arrive at the Applicant’s claimed invention, which is a one-dimensional, decision support scale **that is completely independent of the underlying features of an adaptive speaker verification system**. It is well established in U.S. Patent Law as well as the Manual for Patent Examining Procedure (M.P.E.P.) § 2143.03 that “to establish *prima facie* obviousness of a claimed invention, **all the claim limitations must be taught or suggested by the prior art.**” *In re Royka*, 490 F.2d 981, 180 U.S.P.Q. 580 (C.C.P.A. 1974) (emphasis added).

“All words in a claim must be considered in judging the patentability of that claim against the prior art.” *In re Wilson*, 424 F.2d 1382, 1385, 165 U.S.P.Q. 494, 496 (C.C.P.A. 1970).

Therefore, Claims 23 and 35 overcome the rejection under 35 U.S.C. § 103(a) as being unpatentable over Hamid in view of Campbell et al.

Claim 24 was previously canceled by Applicant and Claims 33, 36 and 46 are currently canceled in this Amendment. Therefore, the rejection of Claims 24, 33, 36 and 46 under 35 U.S.C. § 103(a) as being unpatentable over Hamid in view of Campbell et al. is respectfully believed to be rendered moot.

Claim 44 was rejected under 35 U.S.C. § 103(a) as being unpatentable over Hamid in view of Campbell et al. Since Claim 44 depends from and contains all of the limitations of Claim 35, Claim 44 is felt to distinguish over Hamid in view of Campbell et al. in the same manner as Claim 35. Therefore, Claim 44 overcomes the rejection under 35 U.S.C. § 103(a). If an independent claim is nonobvious under 35 U.S.C. § 103(a), then any claim depending therefrom is nonobvious. *In re Fine*, 837 F.2d 1071, 5 U.S.P.Q.2d 1596 (Fed. Cir. 1988).

Moreover, Claim 44 recites: “wherein the at least one decision criterion defines a single threshold number corresponding to the level of similarity or the level of dissimilarity.” Hamid recites: “...determining if a point in a multidimensional space and having coordinates corresponding substantially to the registration values falls within a multidimensional range determined in dependence upon a predetermined false acceptance rate;....” (Hamid, Column 3, Lines 23-27). Campbell et al. recites: “The algorithm represents an object’s features as small, quantized edge templates, and it represents the object’s geometry with “Hough kernels”. The Hough kernels implement a variant of the generalized Hough transform using simple, 2D image correlation.” (Campbell et al., Page 1, Column 1, Abstract, Lines 12-17).

Therefore, finding a single threshold number corresponding to the level of similarity or the level of dissimilarity provides marked contrast to finding if a point in a multidimensional space and having coordinates corresponding substantially to the registration values falls within a multidimensional range, as found in Hamid, or using a two dimensional algorithm to represent visual objects, as found in Campbell et al. Therefore, one of ordinary skill in the art could not have combined the claimed elements by known methods due to technological difficulties presented by the technology disclosed in Hamid and Campbell et al. as well as the fact that the elements in combination would not achieve the Applicant's claimed Invention. In determining obviousness, the proper analysis is whether the claimed invention would have been obvious to one of ordinary skill in the art after consideration of all the facts. "To reject a claim based on this rationale, U.S. Patent Office personnel must resolve the Graham factual inquiries. Office personnel must then articulate the following: (1) a finding that the prior art included each element claimed, although not necessarily in a single prior art reference, with the only difference between the claimed invention and the prior art being the lack of actual combination of the elements in a single prior art reference; (2) a finding that one of ordinary skill in the art could have combined the elements as claimed by known methods, and that in combination, **each element merely would have performed the same function as it did separately.**" (Federal Register / Volume 72, No. 195 / Wednesday, October 10, 2007 / Notices, Page 57529, "*Examination Guidelines for Determining Obviousness Under 35 U.S.C. § 103 in View of the Supreme Court Decision in KSR International Co. v. Teleflex Inc.*") (emphasis added). In this case, the Applicant's Invention, as claimed, is the only one that defines a single threshold number corresponding to the level of similarity or the level of dissimilarity. This feature would destroy Hamid or Campbell et al. for their intended purposes. If the proposed modification

would render the prior art invention being modified unsatisfactory for its intended purpose, then there is no suggestion or motivation to make the proposed modification. *In re Gordon*, 733 F.2d 900, 221 U.S.P.Q. 1125 (Fed. Cir. 1984). Moreover, "all words in a claim must be considered in judging the patentability of that claim against the prior art." *In re Wilson*, 424 F.2d 1382, 1385, 165 U.S.P.Q. 494, 496 (CCPA 1970).

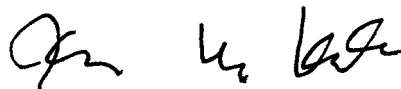
Therefore, Claim 44 overcomes the rejection under 35 U.S.C. § 103(a) as being unpatentable over Hamid in view of Campbell et al.

CONCLUSION

Therefore, it is now believed that all of the pending Claims in the present application are in condition for allowance. Favorable action and allowance of the Claims is therefore respectfully requested. If any issue regarding allowability of any of the pending Claims in the present application could be readily resolved, or if other action could be taken to further advance this application such as an Examiner's Amendment, or if the Examiner should have any questions regarding the present Amendment, it is respectfully requested that the Examiner please telephone the Applicant's undersigned attorney in this regard.

Respectfully submitted,

By:



Kevin M. Kercher, Reg. No. 33,408

Thompson Coburn LLP

One US Bank Plaza

St. Louis, MO 63101-1693

(314) 552-6345

(314) 552-7345 (fax)

Attorney for Applicant

Dated: April 3, 2008

Exhibit A



US005517558A

United States Patent [19]
Schalk

[11] **Patent Number:** **5,517,558**
[45] **Date of Patent:** **May 14, 1996**

[54] **VOICE-CONTROLLED ACCOUNT ACCESS
OVER A TELEPHONE NETWORK**

[75] **Inventor:** Thomas B. Schalk, Dallas, Tex.

[73] **Assignee:** Voice Control Systems, Inc., Dallas,
Tex.

[21] **Appl. No.:** 125,072

[22] **Filed:** Sep. 21, 1993

Related U.S. Application Data

[63] Continuation-in-part of Ser. No. 901,742, Jun. 22, 1992, Pat.
No. 5,297,194, which is a continuation of Ser. No. 523,486,
May 15, 1990, Pat. No. 5,127,043.

[51] **Int. Cl.⁶** G10L 9/08; H04M 1/64;
H04M 1/66

[52] **U.S. Cl.** 379/88; 379/189; 379/199;
381/42; 381/43

[58] **Field of Search** 379/88, 91, 89,
379/188, 189, 199, 95; 381/42, 43

[56] **References Cited**

U.S. PATENT DOCUMENTS

3,896,266	7/1975	Waterbury	381/42 X
4,363,102	12/1982	Holmgren et al.	395/2
4,694,493	9/1987	Sakoe	381/42
4,757,525	7/1988	Matthews et al.	379/89
4,827,518	5/1989	Feustel et al.	381/42
4,850,005	7/1989	Hashimoto	379/51
4,853,953	8/1989	Fujisaki	379/88
4,910,782	3/1990	Watarai	381/42
4,959,855	9/1990	Daudelin	379/213
5,127,043	6/1992	Hunt et al.	379/88
5,181,238	1/1993	Medamana et al.	379/95
5,267,299	11/1993	Nomura	379/88
5,274,695	12/1993	Green	379/88
5,297,194	3/1994	Hunt et al.	379/88

FOREIGN PATENT DOCUMENTS

8910612 11/1989 WIPO.

OTHER PUBLICATIONS

Naik, et al.: Speaker Verification Over Long Distance Tele-
phone Lines, Intl. Conference of Acoustics Speech & Signal
Processing (IEEE), pp. 524-527, May 1989.

Gish: Robust Discrimination in Automatic Speaker Identifi-
cation, 1990 Intl. Conference of Acoustics Speech &
Signal Processing (IEEE), pp.289-292, 1990.

Gong, et al: Text-Independent Speaker Recognition by
Trajectory Space Comparison, 1990 Intl. Conference of
Acoustics, Speech & Signal Processing (IEEE), pp.
285-288, 1990.

Liu, et al.: Study of Line Spectrum Paid Frequencies for
Speaker Recognition, 1990 Intl. Conference of Acoustics,
Speech & Signal Processing (IEEE), pp. 277-280, 1990.

Ren-hua, et al.: A Weighted Distance Measure Based on the
Fine Structure of Feature Space: Application to Speaker
Recognition, 1990 Intl. Conference of Acoustics, Speech &
Signal Processing (IEEE), pp. 273-276, 1990.

Rosenberg, et al.: Sub-Word Unit Talker Verification Using
Hidden Markov Models, 1990 Intl. Conference of Acoustics,
Speed & Signal Processing (IEEE), pp. 269-272, 1990.

(List continued on next page.)

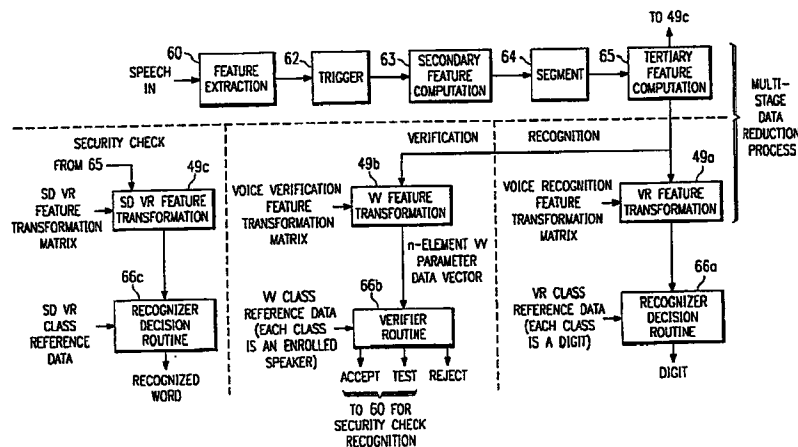
Primary Examiner—Thomas W. Brown

Attorney, Agent, or Firm—David H. Judson

[57] **ABSTRACT**

The present invention describes a system and method for enabling a caller to obtain access to services via a telephone network by entering a spoken first character string having a plurality of digits. Preferably, the method includes the steps of prompting the caller to speak the first character string beginning with a first digit and ending with a last digit thereof, recognizing each spoken digit of the first character string using a speaker-independent voice recognition algorithm, and then following entry of the last digit of the first string, initially verifying the caller's identity using a voice verification algorithm. After initial verification, the caller is again prompted to enter a second character string, which must also be recognized before access is effected.

5 Claims, 3 Drawing Sheets



OTHER PUBLICATIONS

- Bernani, et al.: A Connectionist Approach for Automatic Speaker Identification, 1990 Intl. Conference of Acoustics, Speech & Signal Processing (IEEE), pp. 265-268, 1990.
- Oglesby, et al.: Optimisation of Neural Models for Speaker Identification, 1990 Intl. Conference of Acoustics, Speech & Signal Processing (IEEE), pp. 261-264, 1990.
- Feustal, et al.: Speaker Identity Verification Over Telephone Lines: Where We Are and Where We Are Going, 1989 ICCST, Zurich, Switzerland, pp. 181-182, 1989.
- Velius: Variants of Cepstrum Based Speaker Identity Verification, 1988 Proceedings of ICASSP, pp. 1-4, 1988.
- Doddington: Speaker Recognition-Identifying People by Their Voices, Proceedings of the IEEE, vol. 73, No. 11, pp. 1651-1664, Nov. 1985.
- Gaganelis, et al.: A Novel Approach to Speaker Verification, 1991 Intl. Conference of Acoustics, Speech & Signal Processing (IEEE), pp. 373-376, 1991.
- Bennani, et al.: On The Use of TDNN-Extracted Features Information in Talker Identification, 1991 Intl. Conference of Acoustics, Speech & Signal Processing (IEEE), pp. 385-388, 1991.
- Carey, et al.: A Speaker Verification System Using Alpha-Nets, 1991 Intl. Conference of Acoustics, Speech & Signal Processing (IEEE), pp. 397-400 1991.
- Rose, et al.: Robust Speaker Identification in Noisy Environments Using Noise Adaptive Speaker Models 1991 Intl. Conference of Acoustics, Speech & Signal Processing (IEEE), pp. 401-404, 1991.
- Xu, et al.: The Optimization of Perceptually-Based Features for Speaker Identification, Intl. Conference of Acoustics, Speech & Signal Processing, pp. 520-523 date unknown.
- Perdue, et al.: CONVERSANT® Voice System: Architecture and Applications, AT&T Technical Journal, vol. 65, No. 5, pp. 34-47, Sep./Oct. 1986.
- "Speaker Authentication and Voice Data Entry", Bruno-Beek et al., Conference: 21st Midwest Symposium on Circuits and Systems, Ames, Iowa, 14-15 Aug. 1978, pp. 266-273.

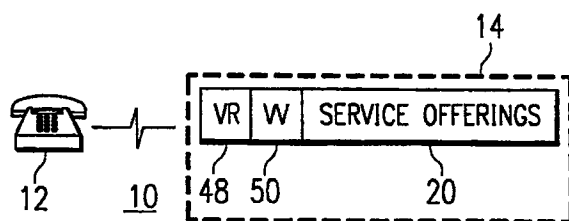


FIG. 1

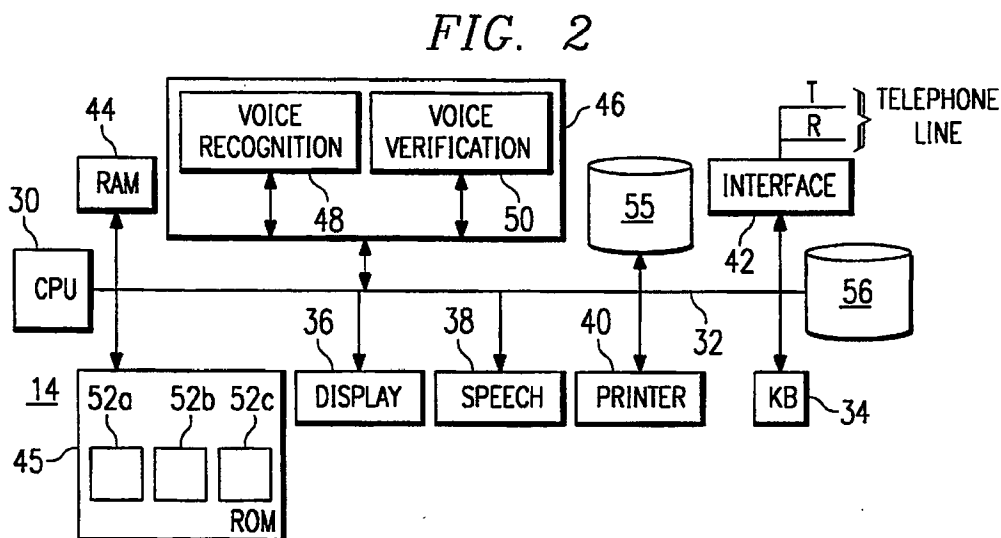
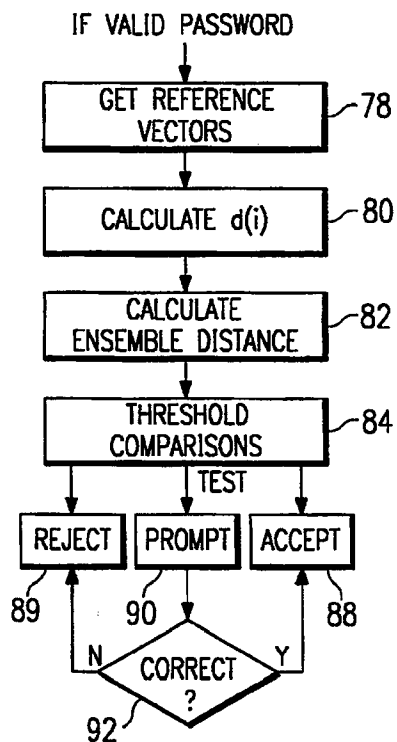


FIG. 4



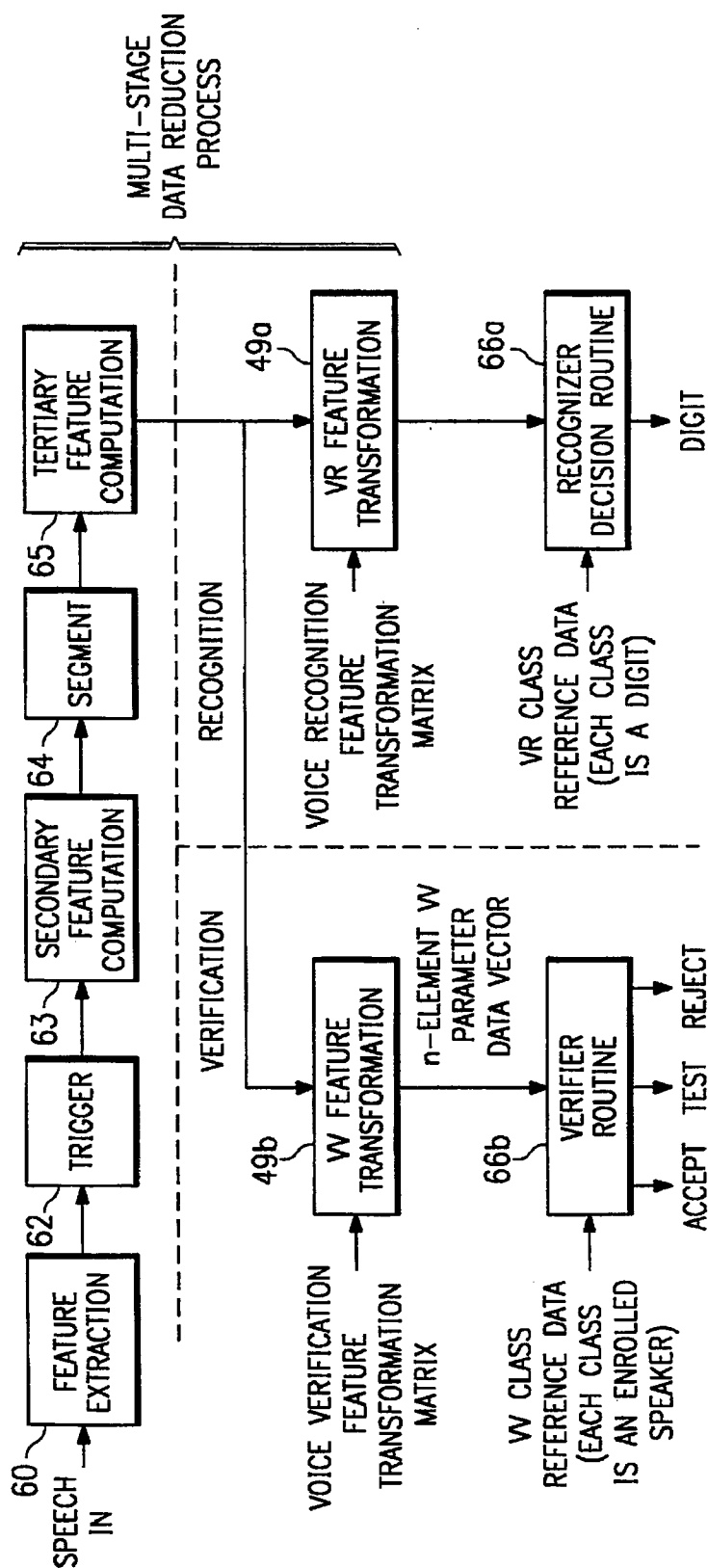


FIG. 3

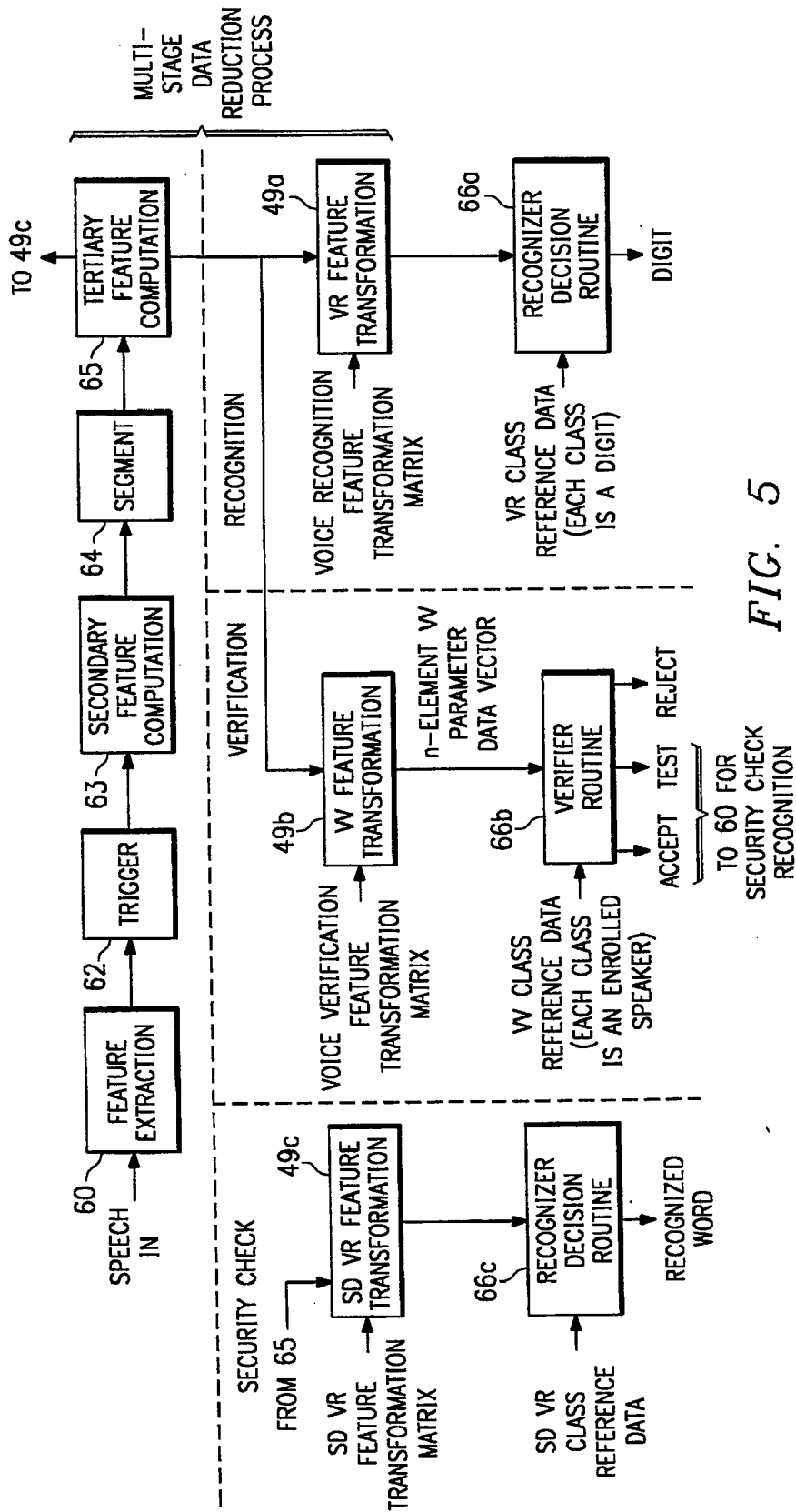


FIG. 5

VOICE-CONTROLLED ACCOUNT ACCESS OVER A TELEPHONE NETWORK

This application is a continuation-in-part of prior application Ser. No. 901,742, filed Jun. 22, 1992, now U.S. Pat. No. 5,297,194, which was a continuation of prior application Ser. No. 07/523,486 filed May 15, 1990, now U.S. Pat. No. 5,127,043.

TECHNICAL FIELD

The present invention relates generally to voice recognition techniques and more specifically to a voice recognition/verification method and system for enabling a caller to obtain access to one or more services via a telephone network.

BACKGROUND OF THE INVENTION

Voice verification is the process of verifying a person's claimed identity by analyzing a sample of that person's voice. This form of security is based on the premise that each person can be uniquely identified by his or her voice. The degree of security afforded by a verification technique depends on how well the verification algorithm discriminates the voice of an authorized user from all unauthorized users.

It would be desirable to use voice verification schemes to verify the identity of a telephone caller. Such schemes, however, have not been successfully implemented. In particular, it has proven difficult to provide cost-effective and accurate voice verification over a telephone network. Generally, this is because the telephone network is a challenging environment that degrades the quality of speech through the introduction of various types of noise and band-limitations. The difficulty in providing telephone-based voice verification is further complicated by the fact that many types of microphones are used in conventional telephone calling stations. These microphones include carbon button handsets, electret handsets and electret speaker phones. Each of these devices possesses unique acoustic properties that affect the way a person's voice may sound over the telephone network.

Given the inherent limitations of the prior art as well as the poor frequency response of the telephone network, it has not been possible to successively integrate a voice recognition and verification system into a telephone network.

BRIEF SUMMARY OF THE INVENTION

It is an object of the present invention to provide a method and system for voice recognition and voice verification over a telephone network.

It is yet another object of the present invention to provide a method and system for enabling a caller to obtain access to one or more services via a telephone network using voice-controlled access techniques.

It is still another object of the invention to provide simultaneous speaker-independent voice recognition and voice verification to facilitate access to services via a band-limited communications channel.

It is another object of the invention to provide a method for verifying the claimed identity of an individual at a telephone to enable the individual to obtain access to services or privileges limited to authorized users.

These and other objects of the invention are provided in a method for enabling a caller to obtain access to services via a telephone network by entering a spoken password having

a plurality of digits. The method begins by prompting the caller to speak the password beginning with a first digit and ending with a last digit thereof. Each spoken digit of the password is then recognized using a speaker-independent voice recognition algorithm. Following entry of the last digit of the password, a determination is made whether the password is valid. If so, the caller's identity is verified using a voice verification algorithm.

This method is implemented according to the invention using a system comprising a digital processor, storage means connected to the digital processor, prompt means controlled by the digital processor for prompting a caller to speak a password beginning with a first digit and ending with a last digit thereof, speech processing means controlled by the digital processor for effecting a multistage data reduction process and generating resultant voice recognition and voice verification parameter data, and voice recognition and verification decision-routines. The storage means includes a read only memory for storing voice recognition feature transformation data and voice recognition class reference data both derived from a first plurality (e.g., 1000) of training speakers over a telephone network. The ROM also stores voice verification feature transformation data derived from a second plurality (e.g., 100-150) of training speakers over a telephone network. The voice recognition feature transformation and class reference data and the voice verification feature transformation data are derived in off-line training procedures. The storage means also includes a database of voice verification class reference data comprising data derived from users authorized to access the services.

The voice recognition routine comprises transformation means that receives the speech feature data generated for each digit and the voice recognition feature transformation data and in response thereto generates voice recognition parameter data for each digit. A digit decision routine receives the voice recognition parameter data and the (digit-relative) voice recognition class reference data and in response thereto generates an output indicating the digit. The voice recognition routine may also include a password validation routine responsive to entry of the last digit of the password for determining if the password is valid.

The voice verification routine is controlled by the digital processor and is responsive to a determination that the password is valid for determining whether the caller is an authorized user. This routine includes transformation means that receives the speech feature data generated for each digit and the voice verification feature transformation data and in response thereto generates voice verification parameter data for each digit. A verifier routine receives the voice verification parameter data and the (speaker-relative) voice verification class reference data and in response thereto generates an output indicating whether the caller is an authorized user.

By way of further background, assume a caller places a call from a conventional calling station telephone to a financial institution or credit card verification company in order to access account information. The caller has previously enrolled in the voice verification database that includes his or her voice verification class reference data. The financial institution includes suitable input/output devices connected to the system (or integrally therewith) to interface signals to and from the telephone line. Once the call setup has been established, the digital processor controls the prompt means to prompt the caller to begin digit-by-digit entry of the caller's preassigned password. The voice recognition algorithm processes each digit and uses a statistical recognition strategy to determine which digit (zero through

nine and "oh") is spoken. After all digits have been recognized, a test is made to determine whether the entered password is valid for the system. If so, the caller is conditionally accepted. In other words, if the password is valid the system "knows" who the caller claims to be and where the account information is stored.

Thereafter, the system performs voice verification on the caller to determine if the entered password has been spoken by a voice previously enrolled in the voice verification reference database and assigned to the entered password. If the verification algorithm establishes a "match," access to the data is provided. If the algorithm substantially matches the voice to the stored version thereof, but not within a predetermined acceptance criterion, the system prompts the caller to input additional personal information (e.g., the caller's social security number or birthdate) to further test the identity of the claimed owner of the password. If the caller cannot provide such information, the system rejects the access inquiry and the call is terminated.

In the preferred embodiment of this invention, even if the verification algorithm establishes a "match" between the entered password and a voice previously enrolled in the voice verification reference database and assigned to the entered password, a further security technique is employed before the caller is provided access to his or her account or to otherwise carry out a transaction. In particular, the caller is prompted to enter some other identifying information which must then be recognized by a preferably speaker-dependent voice recognition algorithm before access is allowed. For example, if the first spoken character string is an "account number," then the additional identifying information may be the caller's social security number or other code. If the first spoken character string was a secret personal identification code, then the additional identifying information may be the caller's account number. In either case, simultaneous recognition and verification is performed on the first character string, at which point the system knows that the caller is who he or she purports to be and that the caller's voice matches (to some acceptable degree) a voice previously enrolled in the voice verification reference database and assigned to the entered character string. According to this preferred embodiment of the invention, the additional security is provided by requiring the caller to further provide the additional identifying information to prevent fraud.

Preferably, the additional identifying information is only valid for a predetermined time period (e.g., one month), and thus the subscriber will contact the service at regular intervals to alter such information. Continuous modification of the additional identifying information further enhances the security of the system.

These objects should be construed to be merely illustrative of some of the more prominent features and applications of the invention. Many other beneficial results can be attained by applying the disclosed invention in a different manner or modifying the invention as will be described. Accordingly, other objects and a fuller understanding of the invention may be had by referring to the following Detailed Description of the preferred embodiment.

BRIEF DESCRIPTION OF THE DRAWINGS

For a more complete understanding of the present invention and the advantages thereof, reference should be made to the following Detailed Description taken in connection with the accompanying drawings in which:

FIG. 1 is a schematic diagram of a telephone network having a calling station connectable to a digital processing system of a service provider such as a financial institution;

FIG. 2 is a schematic diagram of the digital processing system of FIG. 1 for use in providing speaker-independent voice recognition and verification;

FIG. 3 is a block diagram of voice recognition/verification algorithms for use in this invention;

FIG. 4 is a flowchart describing the verifier routine of FIG. 3; and

FIG. 5 is a block diagram of the preferred embodiment of the invention wherein an additional security check is performed before access is allowed to the caller's account.

Similar reference characters refer to similar parts and/or steps throughout the several views of the drawings.

DETAILED DESCRIPTION

FIG. 1 illustrates a block diagram of a conventional telephone network 10 having a calling station 12 connectable to a digital processing system 14 of a financial institution. According to the teachings of the present invention, the digital processing system 14 includes a speaker-independent voice recognition algorithm 48 and an associated voice verification algorithm 50 to facilitate voice-controlled access to one or more services 20 offered by the financial institution. These services include, but are not limited to, account balance inquiry and electronic funds transfer. Moreover, while the following discussion describes the use of voice recognition/verification in the context of accessing information stored in a financial institution, it should be appreciated that the teachings of the invention are not so limited. The invention can be used for numerous other applications such as credit card validation and personal identification validation. Further, it should also be appreciated that the telephone network may include other devices and switching systems conventional in the art. Accordingly, calling station 12 may be connected through a central office or other switching device, such as an access tandem or interexchange carrier switching system, before connection to the service provider.

Referring now to FIG. 2, a block diagram is shown of a digital processing system 14 for use in the present invention to provide the initial step of simultaneous speaker-independent voice recognition and verification. The system, described in U.S. Pat. No. 5,127,043, includes a central processing unit (CPU) 30 for controlling the overall operation of the system. The CPU includes data, address and control buses represented generally by the reference numeral 32. As seen in FIG. 2, the system 14 also includes conventional input/output devices such as a keyboard 34, display terminal 36, speech generator 38 and printer 40. A communications interface 42 (which may be microprocessor-controlled) interfaces the system to the telephone line. Random access memory ("RAM") 44 is connected to the CPU by bus 32 for providing temporary storage of data processed thereby. Read only memory ("ROM") 45 is likewise connected to the digital processor for providing permanent storage of special recognition and verification data as will be described below. Disk storage 46 supports control programs including a voice recognition algorithm 48 and a voice verification algorithm 50 as well as suitable control programs (not shown).

ROM 45 stores voice recognition reference information for use by the voice recognition algorithm 48. This information is of two (2) types: voice recognition feature transformation data 52a and voice recognition class reference data 52b derived from a first plurality of training speakers over a telephone network. In particular, voice recognition

feature transformation data 52a and voice recognition class reference data 52b is derived, in a prior off-line process, from a voice recognition training database (not shown) including "digit" data from a large number of training speakers (e.g., 1000) collected over the telephone network. This training database 52 includes local and long distance data, and significant amounts of data are collected through carbon button handset microphones and electret handset microphones. The voice recognition class reference data 52b includes a representation for each digit word (e.g., "one," "two," etc.) as a "class" sought to be recognized by the voice recognition algorithm 48. For example, the representation of the class for the digit "one" is derived from the data from all of the training speakers who spoke the digit "one."

The voice recognition training database is thus designed to represent the distribution of acoustic characteristics of each digit word across a large population of speakers. The purpose and effect of the analysis performed on this database is to optimize the parameters of a multiple stage data reduction process so as to discover and accurately represent those characteristics of each digit word that differentiate it from each other digit word, regardless of speaker.

ROM 45 also supports voice verification feature transformation data 52c. This data is derived, in a prior off-line process, from a voice verification training database (not shown). In particular, the voice verification training database preferably includes data generated from approximately 100-150 training speakers and is collected over the telephone network. The database includes local and long distance data, and significant amounts of data are collected through carbon button handset microphones and electret handset microphones. Each training speaker is provided with a script containing random digit sequences. The sequences are spoken in a predetermined number (e.g., 5) of separate recording sessions, with the first recording session containing a predetermined number (e.g., 5) of passes of the digits spoken in random order. The subsequent sessions each contain a predetermined number (e.g., 3) of passes of the digits spoken in random order, and each recording session is separated from the previous session by at least one day.

The voice verification training database is thus designed to represent the distribution of acoustic characteristics of each digit word spoken by a particular training speaker across multiple utterances of the digit word by that speaker. The purpose and effect of the analysis performed on this database is to optimize the parameters of a multiple stage data reduction process so as to discover and accurately represent those characteristics of each digit word uttered by each particular training speaker that differentiate it from the same digit word uttered by each other training speaker.

The voice verification technique requires the authorized users of the system (i.e., those persons expected to call over the telephone system to access information) to have previously enrolled in the system. Accordingly, the system 14 also includes a voice verification reference database 55 comprising voice verification class reference data collected from users authorized to access the services. Enrollment is preferably accomplished by having the user speak a ten-digit password five times. For further security, the caller is asked to answer a few factual personal questions that can be answered using digits or words recognizable by the voice recognition algorithm 48. These questions may include, but need not be limited to, the user's social security number, account number or birthdate. Each "class" of the voice verification class reference data represents an authorized user of the system. The class reference data for all authorized users of the system is then stored in the voice verification reference database 55.

The system 14 also includes a transaction database 56 for storing financial and transaction data, such as account balances, credit information and the like. This information is preferably stored at predetermined locations addressed by the caller's password. Thus the password identifies both the caller and the location of the data sought to be accessed.

In operation, as described in U.S. Pat. No. 5,127,043, assume a caller places a call from the calling station 12 to the financial institution in order to access account information. The caller has previously enrolled in the voice verification reference database 55. Once the call setup has been established, the speech generator 38 of the digital processing system 14 prompts the caller to begin digit-by-digit entry of the caller's predetermined password starting with the first digit and ending with the last digit thereof. Prompting of the digits, alternatively, can be effected in any desired manner or sequence. Signals are interfaced to the telephone line by the communications interface 42. As each digit is spoken, the voice recognition algorithm 48 processes the received information and, as will be described below, uses a statistical recognition decision strategy to determine the digit (zero through nine and "oh").

After all digits have been recognized, a test is made to determine whether the entered password is valid for the system. If the outcome of the test is positive, the caller is conditionally accepted because the system "knows" who the caller claims to be and thus where the account information is stored. Thereafter, the system uses the voice verification algorithm 50 to perform voice verification on the caller to determine if the entered password has been spoken by a voice previously enrolled in the database 55 and assigned to the entered password. If the verification algorithm 50 establishes a "match" within predetermined acceptance criteria, access to the data or other system service is allowed (although in the preferred embodiment an additional security check is required as will be described). If the algorithm 50 cannot substantially match the entered voice to a voice stored in the database 55, the system rejects the access inquiry and the call is terminated. If the algorithm 50 substantially matches the entered voice to a voice stored in the database 55, but not within a predetermined acceptance criterion, the system prompts the caller to input additional personal information (e.g., the caller's social security number, account number or other key words) associated with the password to further test the identity of the claimed owner of the password. If the caller cannot provide such additional identifying information, the system rejects the access inquiry and the call is terminated. Correct entry of the requested information enables the caller to gain access to the service.

Referring now to FIG. 3, a block diagram is shown of an embodiment of the voice recognition and verification algorithms 48 and 50 as described in U.S. Pat. No. 5,127,043. As will be seen, algorithms 48 and 50 share the functional blocks set forth in the upper portion of the block diagram. These blocks comprise a speech processing means for carrying out a first tier of a multistage data reduction process. In particular, as speech is input to the system 14, a feature extractor 60 extracts a set of primary features that are computed in real time every 10 milliseconds. The primary features include heuristically-developed time domain features (e.g., zero crossing rates) and frequency domain information such as Fast Fourier Transform ("FFT") coefficients. The output of the feature extractor 60 is a reduced data set (approximately 4,000 data points/utterance instead of the original approximately 8,000 data points/utterance) and is applied to a trigger routine 62 that captures spoken words

using the primary features. The trigger routine is connected to a secondary feature routine 63 for computing "secondary features" from the primary features. The secondary features preferably result from non-linear transformations of the primary features. The output of the routine 63 is connected to phonetic segmentation routine 64. After an utterance is captured and the secondary features are computed, the routine 64 provides automatic phonetic segmentation. To achieve segmentation, the phonetic segmentation routine 64 preferably locates voicing boundaries by determining an optimum state sequence of a two-state Markov process based on a sequence of scalar discriminant function values. The discriminant function values are generated by a two-class Fisher linear transformation of secondary feature vectors. The voicing boundaries are then used as anchor points for subsequent phonetic segmentation.

After the phonetic boundaries are located by the phonetic segmentation routine, the individual phonetic units of the utterance are analyzed and so-called "tertiary features" are computed by a tertiary feature calculation routine 65. These tertiary features preferably comprise information (e.g., means or variances) derived from the secondary features within the phonetic boundaries. The tertiary features are used by both the voice recognition algorithm 48 and the voice verification algorithm 50 as will be described. The output of the routine 65 is a tertiary feature vector of approximately 300 data points/utterance. As can be seen then, the upper portion of FIG. 3 represents the first tier of the multistage data reduction process which significantly reduces the amount of data to be analyzed but still preserves the necessary class separability, whether digit-relative or speaker-relative, necessary to achieve recognition or verification, respectively. The middle portion of FIG. 3 represents a second tier of the data reduction process and, as will be described, comprises the transformation routines 49a and 49b.

To effect speaker-independent voice recognition, the tertiary features are first supplied to the voice recognition linear transformation routine 49a. This routine multiplies the tertiary feature vector by the voice recognition feature transformation data (which is a matrix) 52a to generate a voice recognition parameter data vector for each digit. The output of the transformation routine 49a is then applied to a voice recognition statistical decision routine 66a for comparison with the voice recognition class reference data 52b. The output of the decision routine 66a is a yes/no decision identifying whether the digit is recognized and, if so, which digit is spoken.

Specifically, decision routine 66a evaluates a measure of word similarity for each of the eleven digits (zero through nine, and oh) in the vocabulary. The voice recognition class reference data 52b includes various elements (e.g., acceptance thresholds for each digit class, inverse covariances and mean vectors for each class) used by the decision strategy. For a digit to be declared (as opposed to being rejected), certain acceptance criteria must be met. The acceptance criteria may include, but need not be limited to, the following. The voice recognition algorithm determines the closest match between the class reference data and the voice recognition parameter vector for the digit; this closest match is a so-called "first choice." The next closest match is a "second choice." Each choice has its own matching score. The digit is declared if (1) the matching score of the first choice is below a predetermined threshold, and (2) the difference between the matching score(s) of the first choice and the second choice digits is greater than another predetermined threshold. When all digits of the password have

been recognized, the voice recognition portion of the method is complete.

To effect voice verification, the tertiary features are also supplied to a linear transformation routine 49b that multiplies each tertiary feature vector by the voice verification feature transformation data (which is a matrix). The output of the routine 49b is an N_p -element vector p of voice verification parameter data for each digit of the password, with N_p preferably approximately equal to 25. The voice verification parameter data vector p is then input to a verifier routine 66b which also receives the voice verification class reference data 52c for the caller. Specifically, the voice verification class reference data is provided from the voice verification reference database 55. As noted above, the address in the database 55 of the caller's voice verification class reference data is defined by the caller's password derived by the voice recognition algorithm 48.

Verifier routine 66b generates one of three different outputs: ACCEPT, REJECT and TEST. An ACCEPT output may authorize the caller to access data from the transaction database 56. The REJECT output is provided if the verifier disputes the purported identity of the caller. The TEST output initiates the prompting step wherein additional follow-up questions are asked to verify the caller's identity.

Referring now to FIG. 4, a flowchart is shown of verifier routine 66b of FIG. 3. By way of background, the routine begins after the determination, preferably by the voice recognition algorithm 48, that the password is valid. Although in the preferred embodiment each voice verification parameter vector is generated as each digit is recognized, it is equally possible to refrain from generating the voice verification parameter vectors until after a test is performed to determine whether the password is valid.

The verifier routine begins at step 78. In particular, the N_p -element voice verification parameter vectors for each digit of the spoken password are compared with the previously-generated voice verification class reference data vectors stored in the voice verification reference database 55. First, a weighted Euclidean distance $d(i)$ is computed for each digit at step 80:

$$d(i) = \left[\sum_{j=1}^{N_p} w_j (p(i, j) - pr(i, j))^2 \right]^{1/2}$$

where:

$p(i, j)$ is the j th component of the length- N_p vector generated from the i th digit in the length- N_d current password entry sequence,

$pr(i, j)$ is the j th component of the reference vector of the i th digit for the alleged enrolled caller,

w_j is a constant weighting vector, precalculated to yield optimum system performance, and

$d(i)$ is the resultant weighted Euclidean distance measure for the i th digit in the current password entry sequence.

The distance vector d is then sorted in ascending order:

$$d(i), \dots, d(N_d) = \min_{i=1}^{N_d} (d(i)), \dots, \max_{i=1}^{N_d} (d(i))$$

An ensemble distance is then calculated at step 82 as a weighted combination of these sorted distances:

$$D = \sum_{i=1}^{Nd} w_2(i) d(i)$$

where:

d is the sorted distance vector

w_2 is another constant weighting vector, precalculated to yield optimum system performance, and

D is the resultant ensemble distance measure for the entire current password entry sequence, with respect to the alleged enrolled caller.

At step 84, the ensemble distance is compared to two (2) acceptance thresholds, an upper threshold and a lower threshold. If the ensemble distance is below the lower acceptance threshold, the test is positive and the caller gains immediate access to the requested service. This is the ACCEPT output 88. If the distance is greater than the upper threshold, the caller's access to the service is denied and the method terminates. This corresponds to the REJECT output 89. If the outcome of the test 84 is between the upper and lower thresholds, the method continues at step 90 by prompting the caller to answer one or more factual questions uniquely associated with the password. This is the TEST output. For example, the caller is requested to speak his/her social security number or his/her account number. Alternatively, the caller can be prompted to enter such identifying information manually through the telephone keypad or by pulling a credit card or the like through a card reader. Of course, the nature and scope of the personal information requested by the system depends entirely on the system operator and the degree of security sought by the caller and operator. A test is then performed at step 92 to determine if the question(s) have been correctly answered. If the outcome of the test is positive, the caller again gains access to the requested service. If the outcome of the test at step 92 is negative, access is denied and the method terminates.

Accordingly, the above described system provides a voice recognition/verification system and method having several advantages over prior art telephone-based data access schemes. The problems inherent in the limited frequency response environment of a telephone network are ameliorated through the use of a speaker-independent voice recognition system and a voice verification algorithm. The voice verification algorithm is "trained" by a voice verification training database that includes speaker classifications as opposed to word classifications. Moreover, the verification algorithm uses tertiary features and voice verification feature transformation parameters to calculate a preferably 25-element vector for each spoken digit of the entered password. These vectors are then compared with voice verification class reference data (for the caller) and a weighted Euclidean distance is calculated for each digit. An ensemble distance for the entire password is then computed and compared to two acceptance thresholds to determine if the caller's voice matches his or her previously stored voice templates. Callers who "almost match" must get through an additional level of security before access to the data or service is authorized.

The digital processing system may be, but is not limited to, a IBM AT personal computer which is connected to a local area network for storing and accessing verification reference data. For telephone-based applications requiring confidential access to information, the system 14 has numerous applications. By way of example only, voice verification over the telephone network has significant potential for eliminating calling card fraud. In addition, banks and other financial institutions can provide more security to telephone-

based account access systems. Presently, banking systems use personal identification numbers of "PIN" digits entered via the telephone keypad to determine eligibility for system entry. Voice verification as well as PIN digits may be employed to determine if a caller is authorized for access to account information. Other uses for the system described above include credit information access, long distance telephone network access, and electronic funds transfer. Because the voice verification operates in conjunction with voice recognition, rotary telephone users are also able to use any automated application employing the system.

In the preferred embodiment, it is desirable to provide additional security to the system. This embodiment is shown in FIG. 5, which is a modification to the system shown in FIG. 3. In this embodiment, again assume a caller places a call from a conventional calling station telephone to a financial institution or credit card verification company in order to access account information. The caller has previously enrolled in the voice verification database that includes his or her voice verification class reference data. The financial institution includes suitable input/output devices connected to the system (or integrally therewith) to interface signals to and from the telephone line. Once the call setup has been established, the digital processor controls the prompt means to prompt the caller to begin entry of a first character string. For exemplary purposes, it is assumed that the first character string is an account number. Of course, the first character string may be a secret password known only to caller. The voice recognition algorithm processes each character (in either a discrete or continuous fashion) and uses the statistical recognition strategy to determine which character is spoken as previously described with respect to FIG. 3. After all characters of the first character string have been recognized, a test may be made to determine whether the entered string is valid for the system. This step may be omitted. If the entered string is valid, the caller is conditionally accepted.

Thereafter, as previously described the system performs voice verification on the caller to determine if the entered character string has been spoken by a voice previously enrolled in the voice verification reference database and assigned to the entered password. If the verification algorithm establishes a "match," the system knows that the caller is who he or she purports to be and that the caller's voice matches (to some acceptable degree) a voice previously enrolled in the voice verification reference database and assigned to the entered character string. By "match" it is meant that the result of the verifier routine is either an ACCEPT or TEST output. In either case, however, an additional security check is performed (although it may be desirable to perform the additional security check only for the TEST output). Like the FIG. 3 embodiment, the system prompts the caller to input additional information. If the first character string was an account number, then the additional information may be caller's social security number, birthdate, or other keywords. If the first character string was itself a secret password, then the additional information might be the caller's account number. The additional security level, in either case, allows the system to further test the identity of the claimed owner of the first character string, even where the original verifier output was ACCEPT.

As seen in FIG. 5, after the caller is again prompted to enter the additional identifying information (which will be referred to hereinafter as the second character string), the string is processed again by the multi-stage data reduction process (elements 60, 62, 63, 64 and 65). At this point, the second character string is applied to a speaker-dependent

voice recognition feature transformation 49c, which receives as its other input a speaker-dependent voice recognition feature transformation matrix as previously described. The output of the transformation 49c is supplied to a recognizer decision routine 66c, which receives as its other input speaker-dependent voice recognition class reference data. The output of the recognizer decision routine is a speaker-dependent word that the system must accept as the second character string before the transaction is effected. If the caller cannot provide the second character string or, if the caller provides an unrecognizable second character string associated with the first character string, then the system rejects the access inquiry and the call is terminated.

Thus according to this embodiment, even if the verification algorithm establishes a "match" between the entered password and a voice previously enrolled in the voice verification reference database and assigned to the entered password, a further security technique is employed before the caller is provided access to his or her account or to otherwise carry out a transaction. In particular, the caller is prompted to enter some other identifying information (preferably a secret password) which must then be recognized by a preferably speaker-dependent voice recognition algorithm before access is allowed. Thus simultaneous recognition and verification is performed on a first character string, at which point the system knows that the caller is who he or she purports to be and that the caller's voice matches (to some acceptable degree) a voice previously enrolled in the voice verification reference database and assigned to the entered first character string. Additional security is then provided by requiring the caller to further provide a second character string which must be recognized before the transaction is effected.

Preferably, the system requires that the authorized callers change their identifying information on a periodic basis (e.g., monthly). Thus a subscriber's additional identifying information will only be valid for a predetermined time period.

It should be appreciated by those skilled in the art that the specific embodiments disclosed above may be readily utilized as a basis for modifying or designing other structures for carrying out the same purposes of the present invention. For example, the voice recognition algorithm 48 could alternatively be speaker-dependent instead of speaker-independent as described in the preferred embodiment. It should also be realized by those skilled in the art that such equivalent constructions do not depart from the spirit and scope of the invention as set forth in the appended claims.

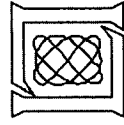
What is claimed is:

1. A method for enabling a caller to obtain access to one or more services via a telephone network by speaking first and second character strings each having a plurality of characters, comprising the steps of:

- (a) prompting the caller to speak the first character string beginning with a first character and ending with a last character thereof;
- (b) generating speech feature data for each spoken character of the first character string;
- (c) applying the speech feature data and voice recognition feature transformation data to a voice recognition feature transformation to generate a first set of parameters for each spoken character of the first character string, the first set of parameters for use in a voice recognition system;

- (d) applying the speech feature data and voice verification feature transformation data to a voice verification feature transformation to generate a second set of parameters for each spoken character of the first character string, the second set of parameters for use in a voice verification system;
 - (e) recognizing the first character string using the first set of parameters;
 - (f) initially verifying the caller's identify using the second set of parameters generated for the first character string; and
 - (g) repeating steps (a)-(c) and (e) using the second character string instead of the first character string to confirm the caller's identity.
2. The method as described in claim 1 wherein the second character string confirms the caller's identity only during a predetermined time period.
3. A method for enabling a caller to obtain access to one or more services via a telephone network by speaking first and second character strings each having one or more characters, comprising the steps of:
- (a) prompting the caller to speak the first character string beginning with a first character and ending with a last character thereof;
 - (b) generating speech feature data for each spoken character of the first character string;
 - (c) applying the speech feature data of the first character string and voice recognition feature transformation data to a voice recognition feature transformation to generate a first set of parameters for each spoken character of the first character string, the first set of parameters for use in a voice recognition system;
 - (d) applying the speech feature data and voice verification feature transformation data to a voice verification feature transformation to generate a second set of parameters for each spoken character of the first character string, the second set of parameters for use in a voice verification system;
 - (e) recognizing the first character string using the first set of parameters of the first character string;
 - (f) initially verifying the caller's identify using the second set of parameters generated for the first character string;
 - (g) prompting the caller to enter the second character string beginning with a first character and ending with a last character thereof;
 - (h) generating speech feature data for each spoken character of the second character string;
 - (i) applying the speech feature data of the second character string and voice recognition feature transformation data to a voice recognition feature transformation to generate a first set of parameters for each spoken character of the second character string, the first set of parameters of the second character string for use in a voice recognition system; and
 - (j) recognizing the second character string using the first set of parameters of the second character string.
4. The method of claim 3 further including the step of determining if the recognized second character string is a password associated with the caller verified in step (f).
5. The method as described in claim 3 further including the step of periodically changing the second character string for confirming the identity of the caller.

Exhibit B



Automatic Speaker Recognition

Recent Progress, Current Applications, and Future Trends

Douglas A. Reynolds, PhD

Senior Member of Technical Staff

M.I.T. Lincoln Laboratory

Larry P. Heck, PhD

Manager, Speaker Verification R&D

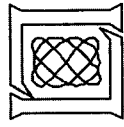
Nuance Communications

**Presented at the AAAS 2000 Meeting
Humans, Computers and Speech Symposium
19 February 2000**

This work was sponsored by the Department of Defense under Air Force contract F19628-95-C-0002. Opinions, interpretations, conclusions, and recommendations are those of the authors and are not necessarily endorsed by the United States Air Force.

Nuance Communications

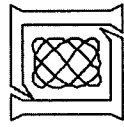
MIT Lincoln Laboratory



Outline



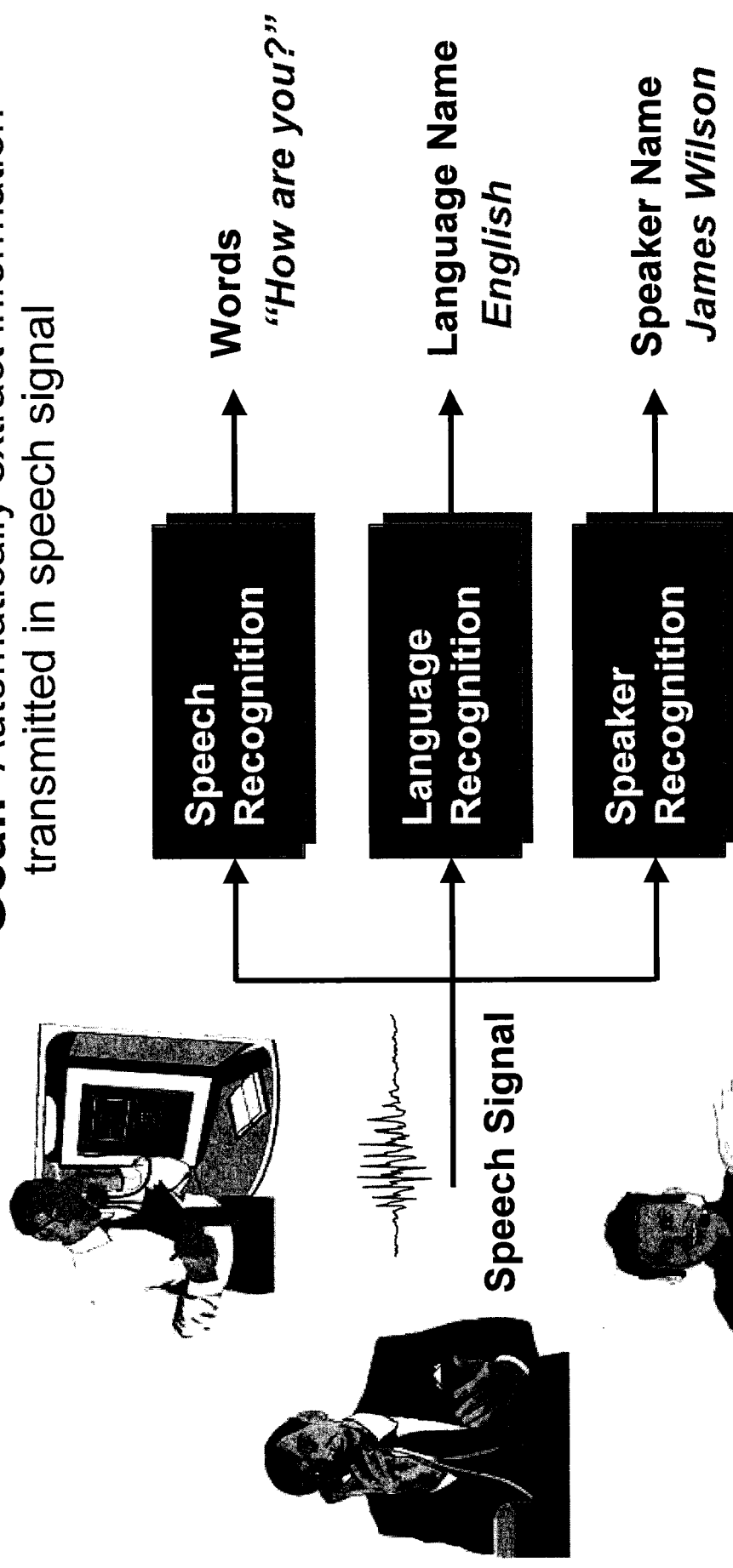
- Introduction (Reynolds)
- General theory (Reynolds)
- Performance (Heck)
- Applications (Heck)
- Conclusions and future directions (Heck)

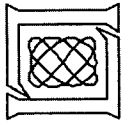


Extracting Information from Speech



Goal: Automatically extract information transmitted in speech signal

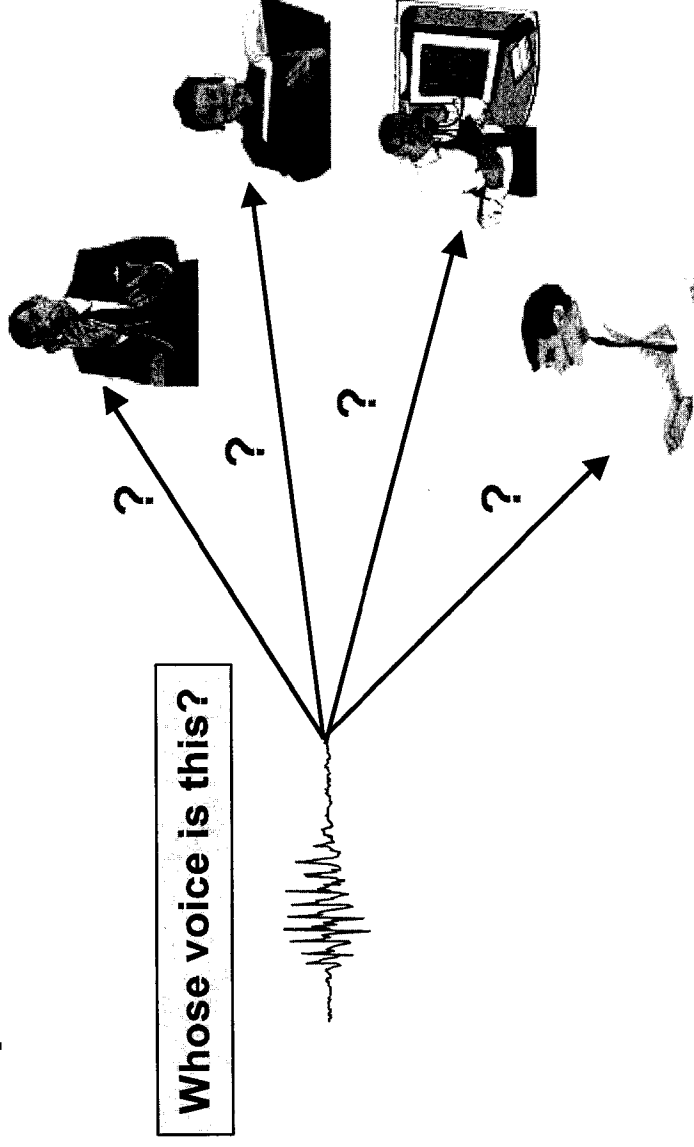


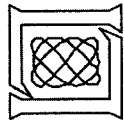


Introduction Identification



- Determines who is talking from set of known voices
- No identity claim from user (many to one mapping)
- Often assumed that unknown voice must come from set of known speakers - referred to as closed-set identification



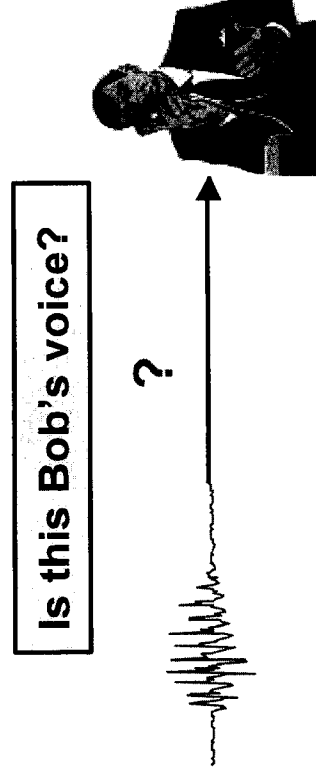


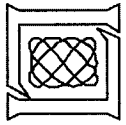
Introduction

Verification/Authentication/Detection



- Determine whether person is who he/she claims to be
- User makes identity claim: one to one mapping
- Unknown voice could come from large set of unknown speakers - referred to as open-set verification
- Adding “none-of-the-above” option to closed-set identification gives open-set identification





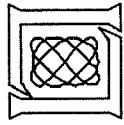
Introduction

Speech Modalities



Application dictates different speech modalities:

- **Text-dependent recognition**
 - Recognition system knows text spoken by person
 - Examples: fixed phrase, prompted phrase
 - Used for applications with strong control over user input
 - Knowledge of spoken text can improve system performance
- **Text-independent recognition**
 - Recognition system does not know text spoken by person
 - Examples: User selected phrase, conversational speech
 - Used for applications with less control over user input
 - More flexible system but also more difficult problem
 - Speech recognition can provide knowledge of spoken text



Introduction

Voice as a Biometric

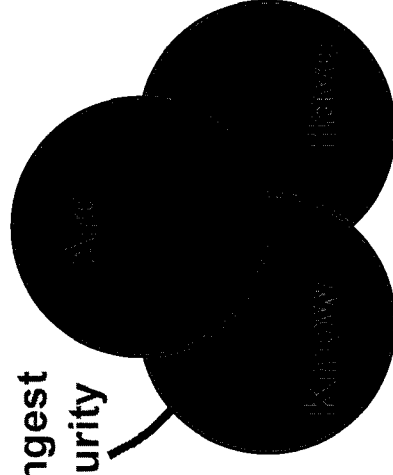


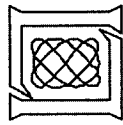
- **Biometric:** a human generated signal or attribute for authenticating a person's identity
- **Voice is a popular biometric:**
 - natural signal to produce
 - does not require a specialized input device
 - ubiquitous: telephones and microphone equipped PC

- **Voice biometric with other forms of security**

–	Something you have - e.g., badge
–	Something you know - e.g., password
–	Something you are - e.g., voice

Strongest
security

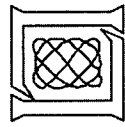




Outline

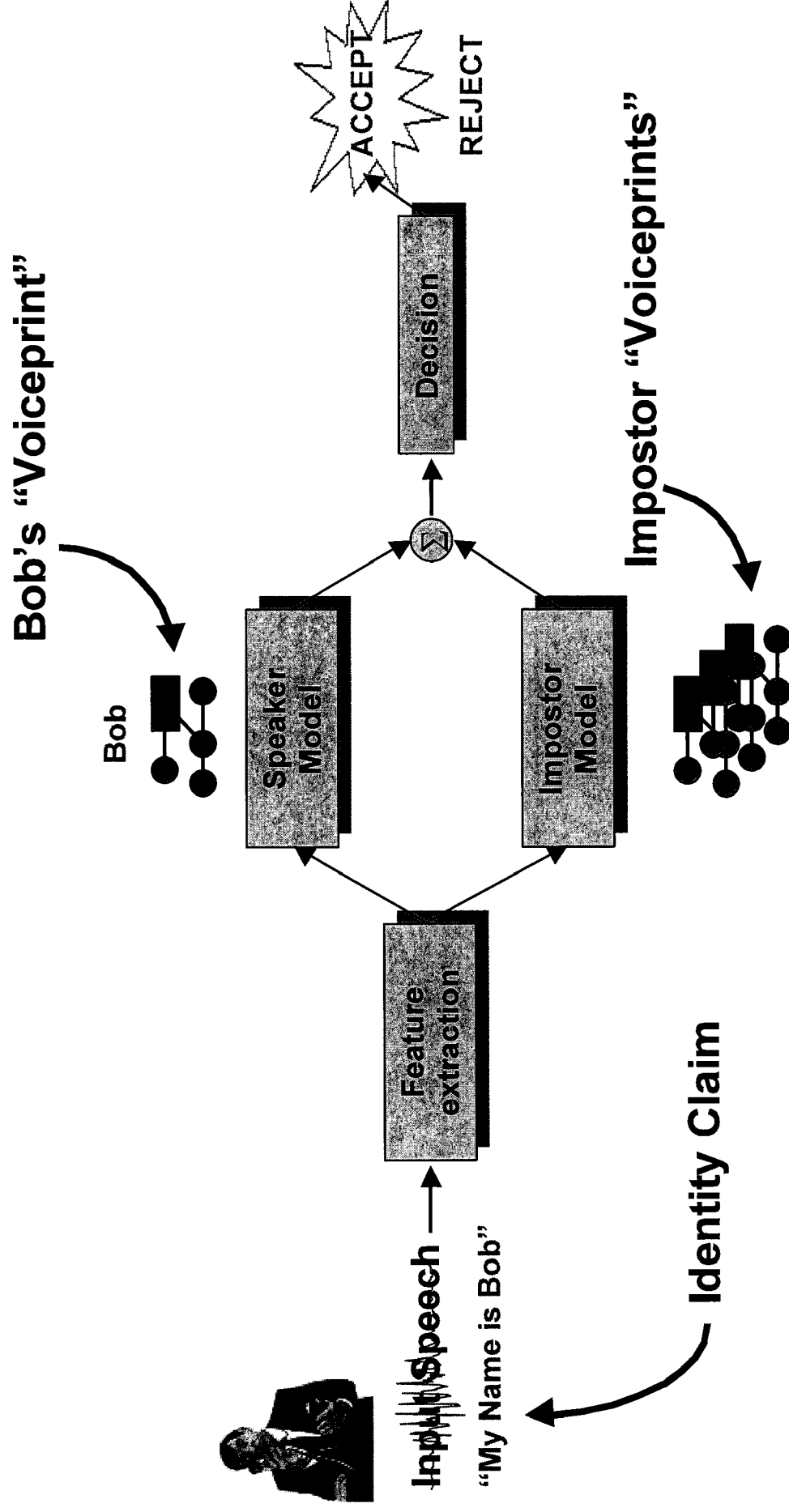


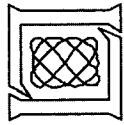
- Introduction
- General theory
- Performance
- Applications
- Conclusions and future directions



General Theory

Components of Speaker Verification System





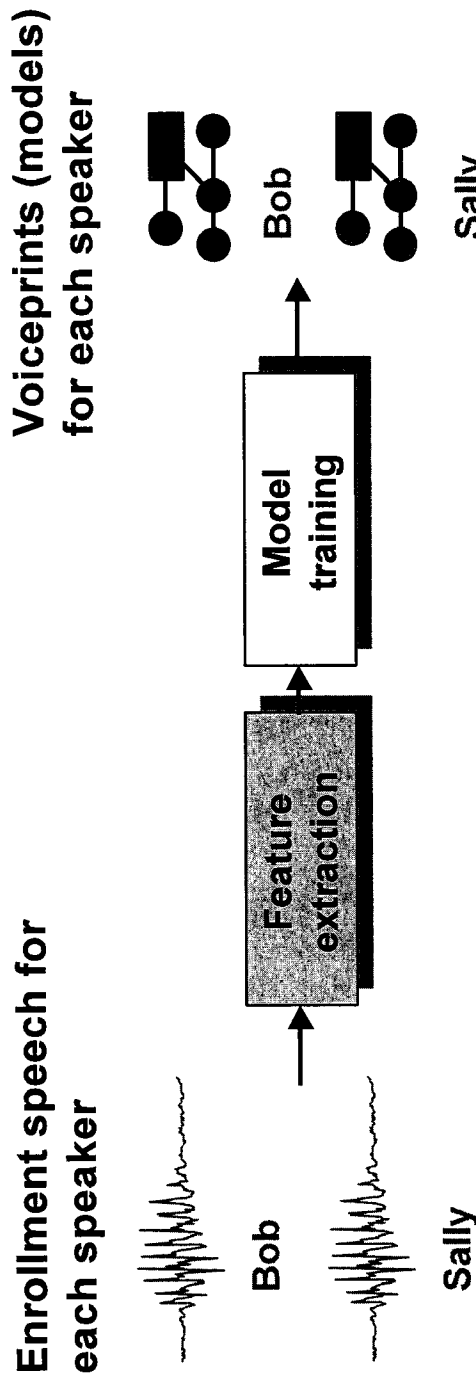
General Theory

Phases of Speaker Verification System

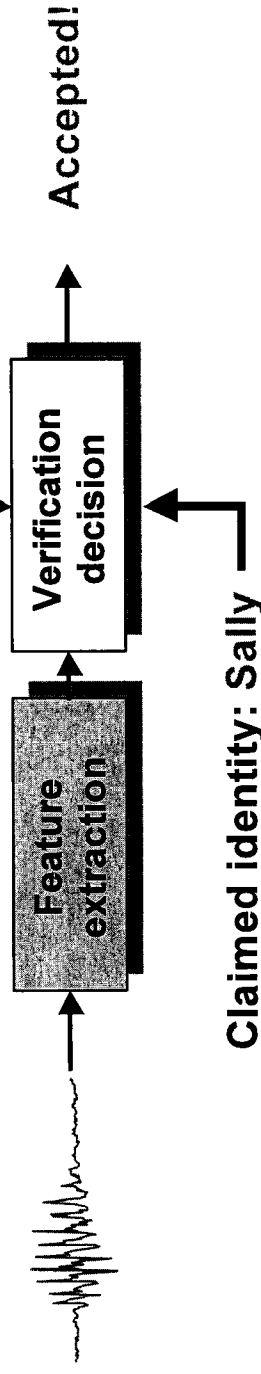


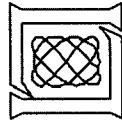
Two distinct phases to any speaker verification system

Enrollment Phase



Verification Phase



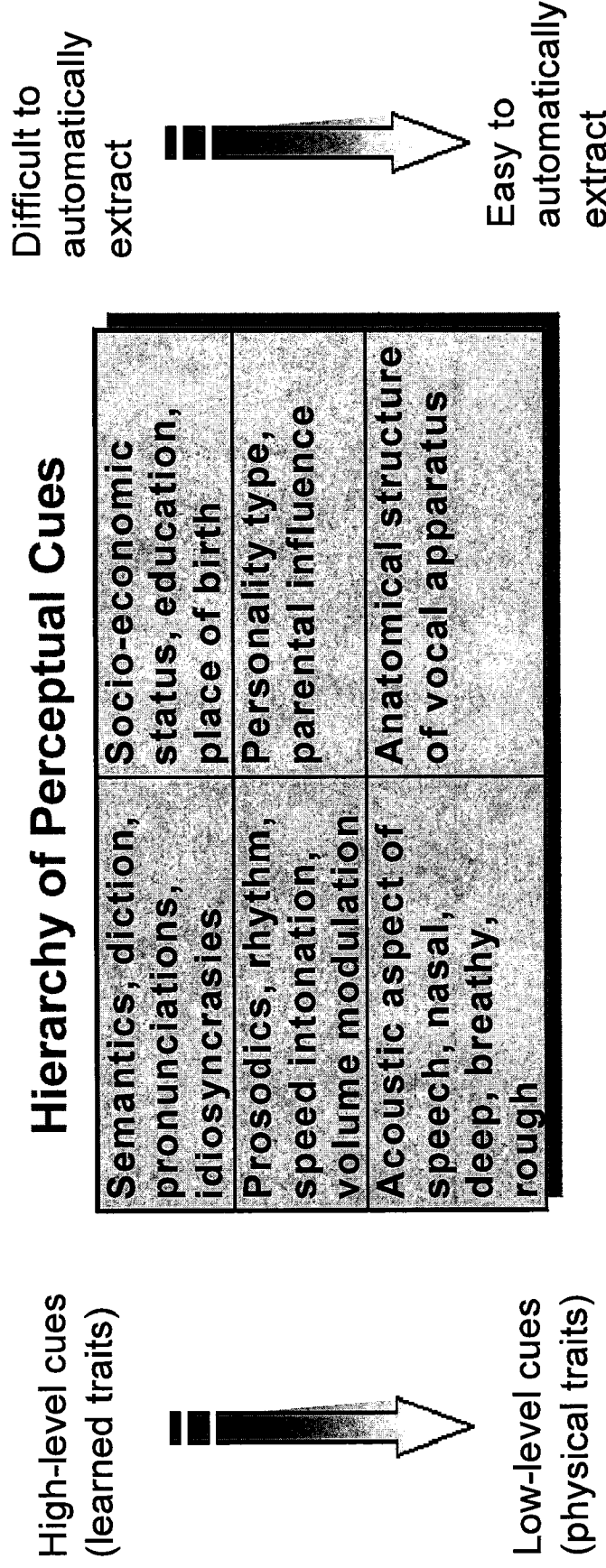


General Theory

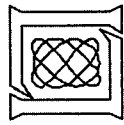
Features for Speaker Recognition



- Humans use several levels of perceptual cues for speaker recognition



- There are no exclusive speaker identity cues
- Low-level acoustic cues most applicable for automatic systems



General Theory

Features for Speaker Recognition



- Desirable attributes of features for an automatic system (Wolf '72)

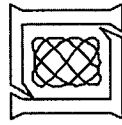
Practical

Robust

Secure

- Occur naturally and frequently in speech
- Easily measurable
- Not change over time or be affected by speaker's health
- Not be affected by reasonable background noise nor depend on specific transmission characteristics
- Not be subject to mimicry

- No feature has all these attributes
- Features derived from spectrum of speech have proven to be the most effective in automatic systems



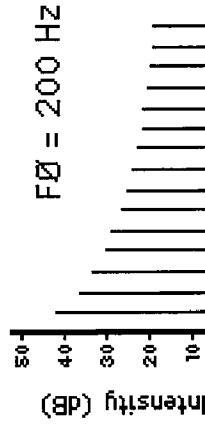
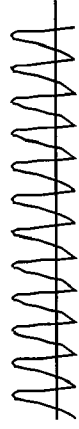
General Theory

Speech Production



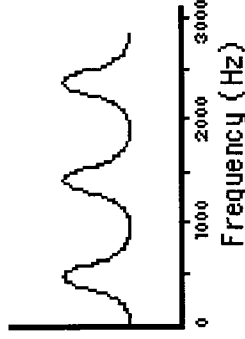
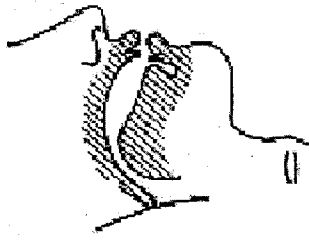
- **Speech production model: source-filter interaction**
 - Anatomical structure (vocal tract/glottis) conveyed in speech spectrum

Glottal pulses



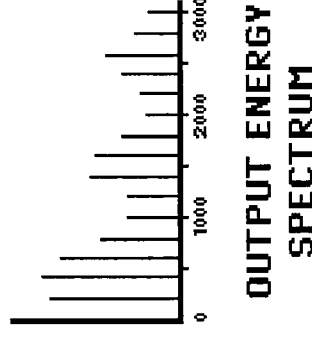
SOURCE SPECTRUM

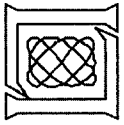
Vocal tract



FILTER FUNCTION

Speech signal



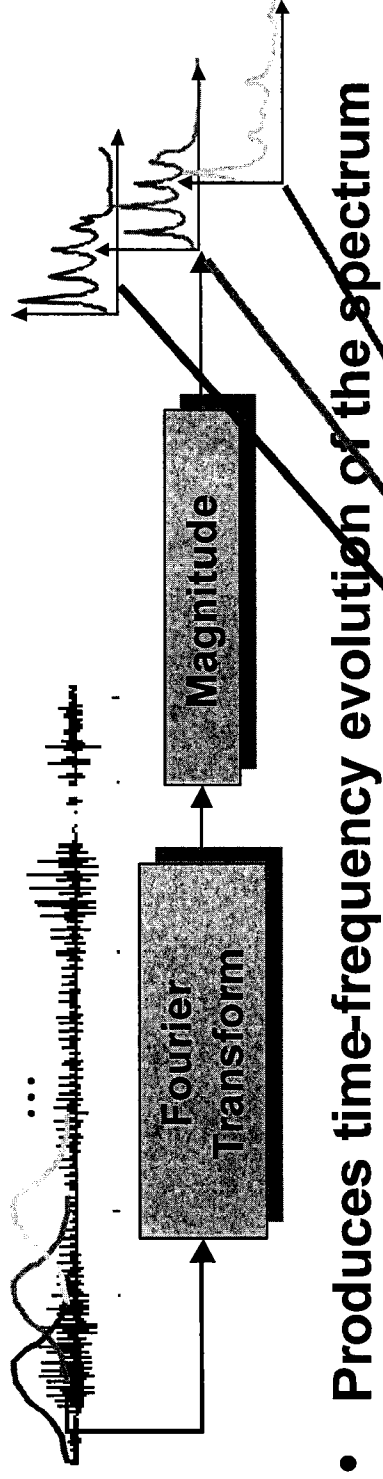


General Theory

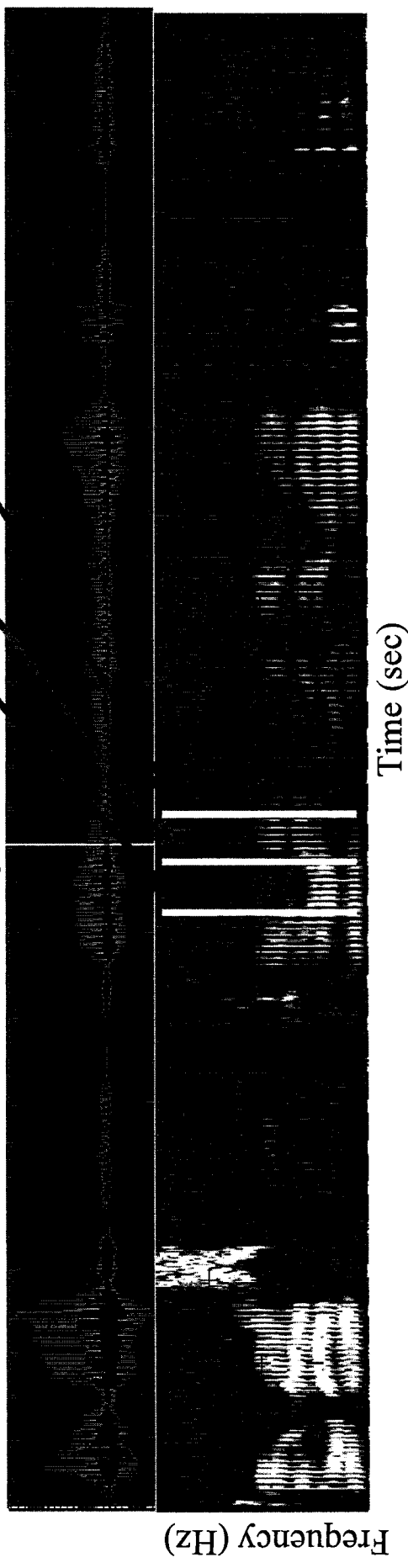
Features for Speaker Recognition

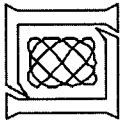


- **Speech is a continuous evolution of the vocal tract**
 - Need to extract time series of spectra
 - Use a sliding window - 20 ms window, 10 ms shift

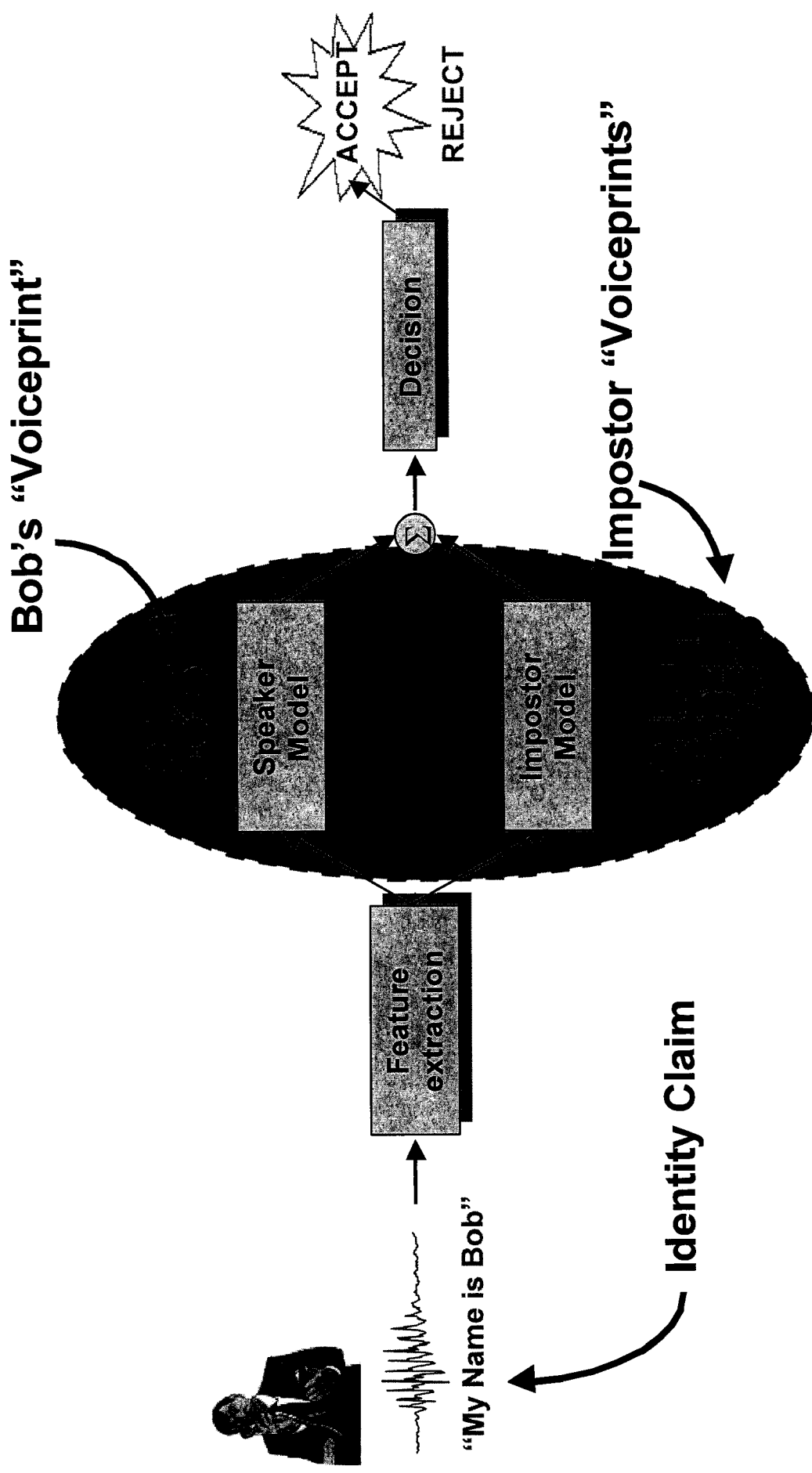


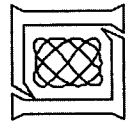
- **Produces time-frequency evolution of the spectrum**





General Theory Speaker Models





General Theory

Speaker Models

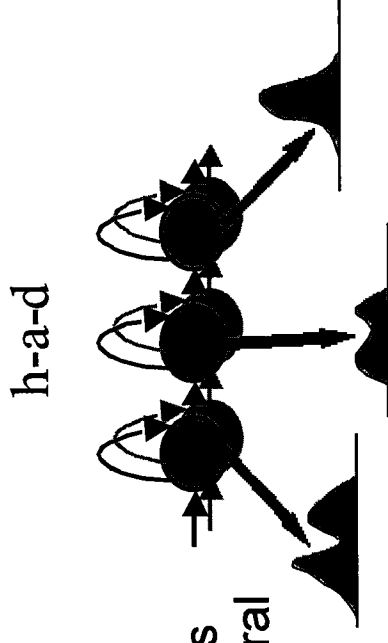


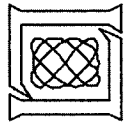
- **Speaker models (voiceprints) represent voice biometric in compact and generalizable form**
- **Modern speaker verification systems use Hidden Markov Models (HMMs)**

- HMMs are statistical models of how a speaker produces sounds

- HMMs represent underlying statistical variations in the speech state (e.g., phoneme) and temporal changes of speech between the states.

- Fast training algorithms (EM) exist for HMMs with guaranteed convergence properties.



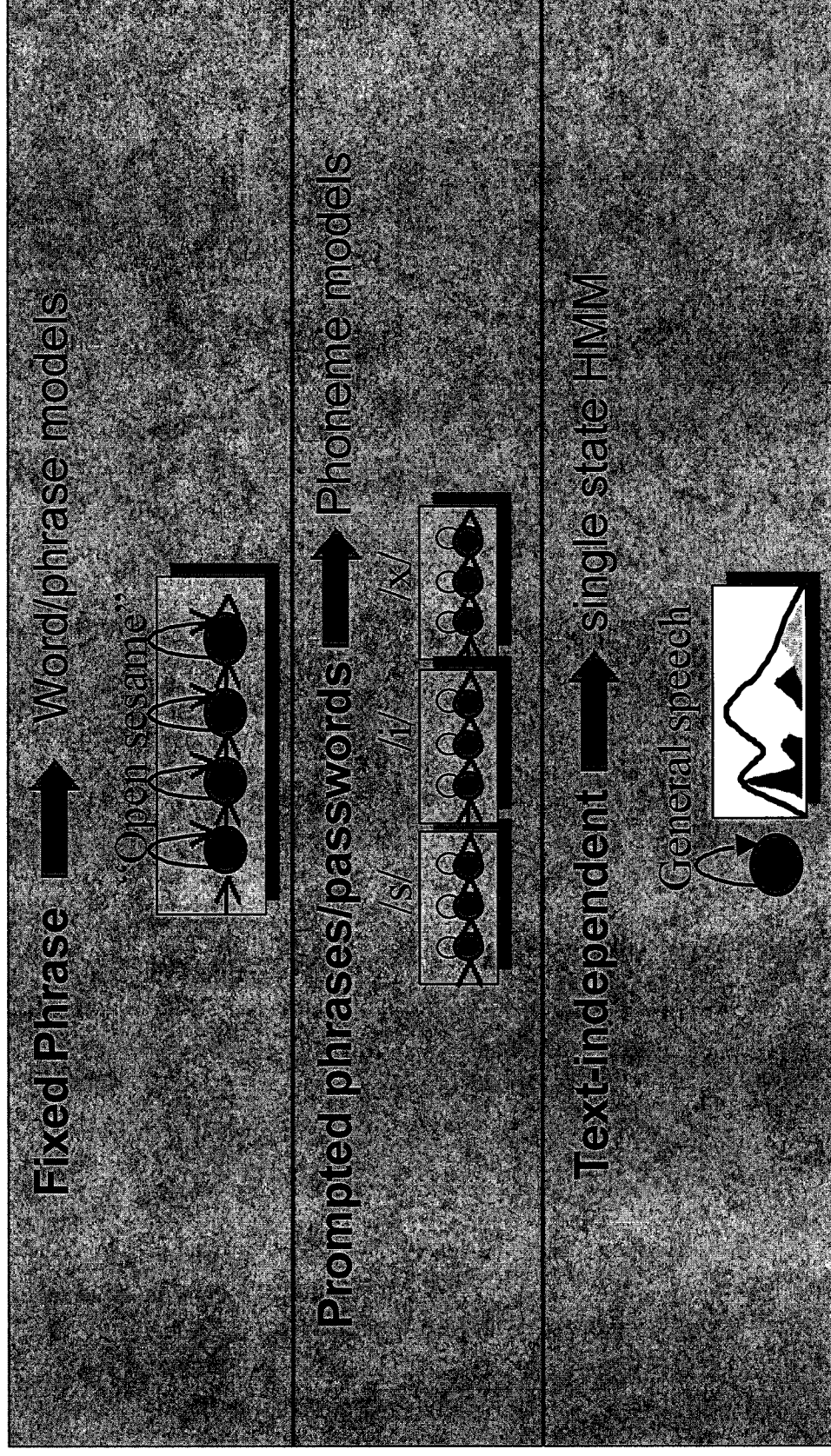


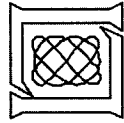
General Theory

Speaker Models



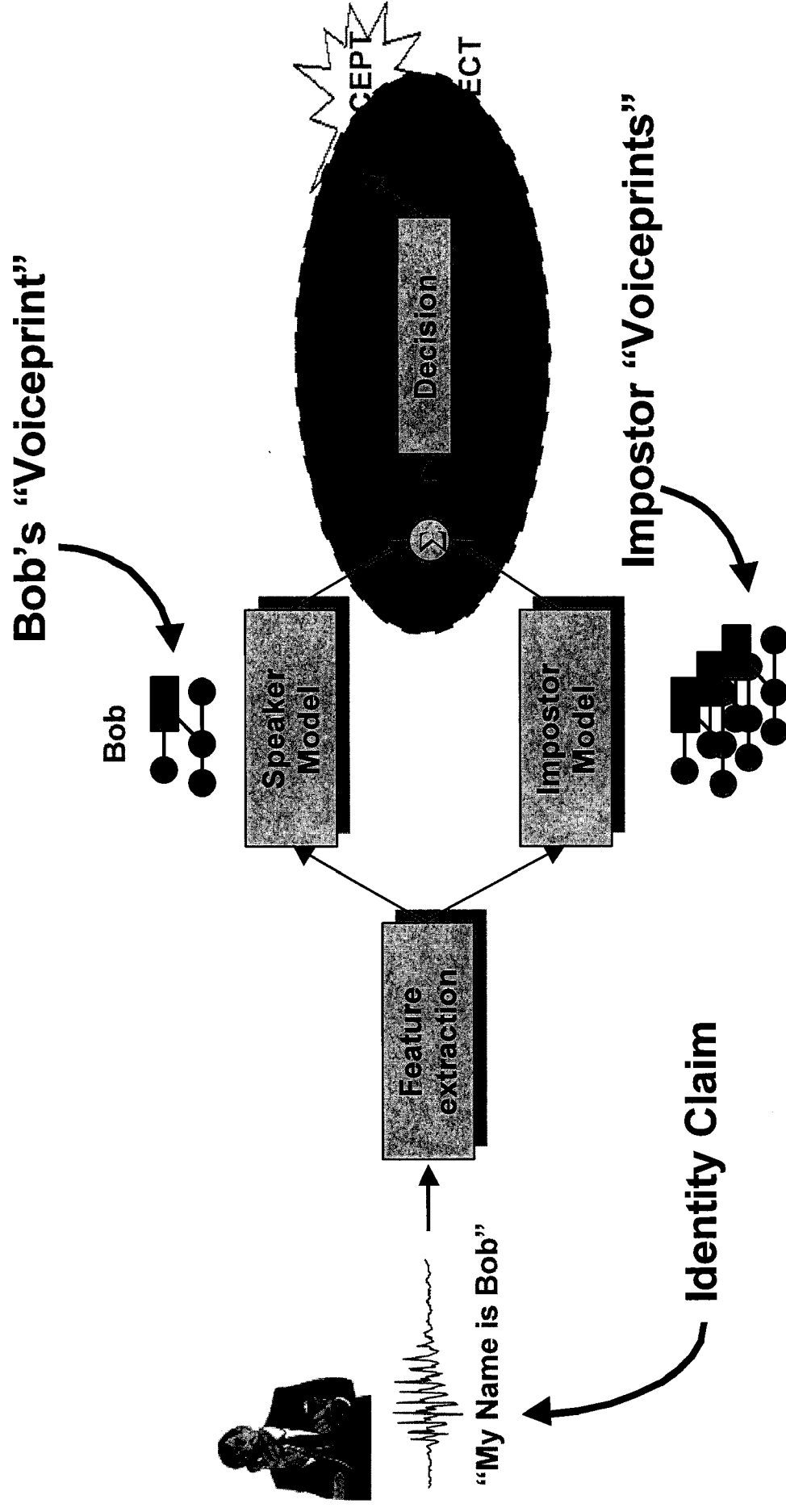
Form of HMM depends on the application

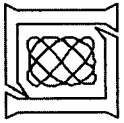




General Theory

Verification Decision





General Theory

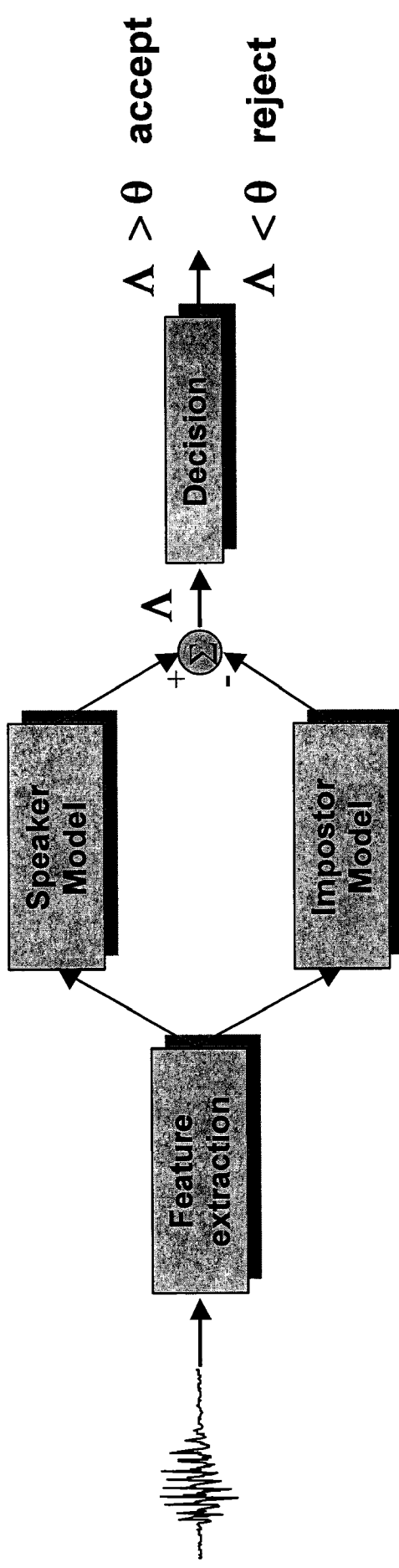
Verification Decision

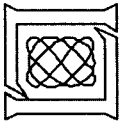


Verification decision approaches have roots in signal detection theory

- **2-class Hypothesis test:**
 - H0: the speaker is an impostor
 - H1: the speaker is indeed the claimed speaker.
- **Statistic computed on test utterance **S** as likelihood ratio:**

$$\Lambda = \log \frac{\text{Likelihood } \mathbf{S} \text{ came from speaker HMM}}{\text{Likelihood } \mathbf{S} \text{ did not come from speaker HMM}}$$

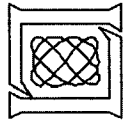




Outline



- Introduction
- General theory
- Performance
- Applications
- Conclusions and future directions



Verification Performance

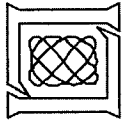
Evaluating Speaker Verification Systems



- There are many factors to consider in evaluating speaker verification systems

Speech quality	<ul style="list-style-type: none">– Channel and microphone characteristics– Noise level and type– Variability between enrollment and verification speech
Speech modality	<ul style="list-style-type: none">– Fixed/prompted/user-selected phrases– Free text
Speech duration	<ul style="list-style-type: none">– Duration and number of sessions of enrollment and verification speech
Speaker population	<ul style="list-style-type: none">– Size and composition

The evaluation data and design should match the target application domain of interest

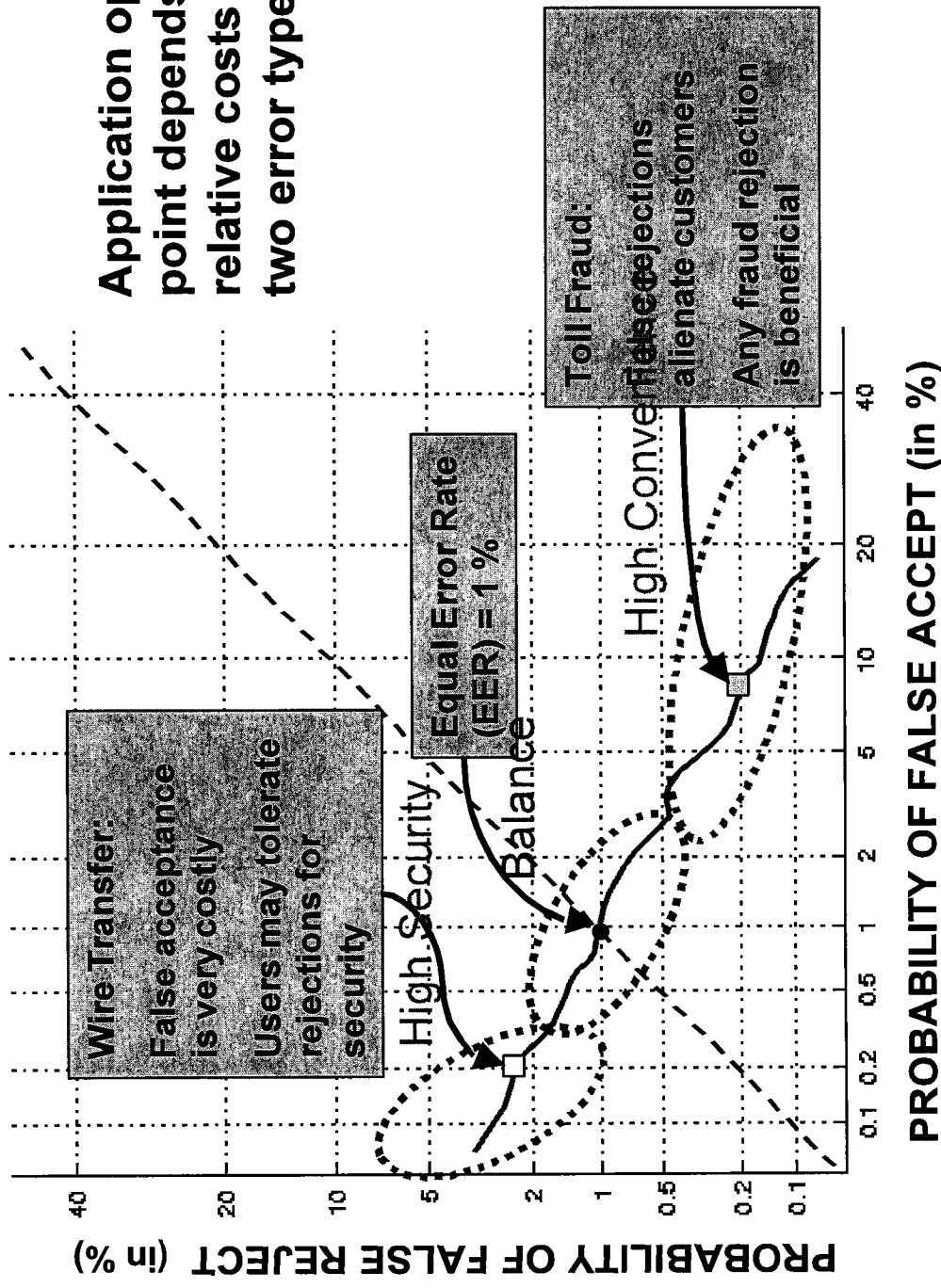


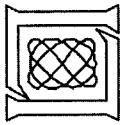
Verification Performance

Evaluating Speaker Verification Systems



Example Performance Curve : Detection Error Tradeoff (DET) Curve





Verification Performance

NIST Speaker Verification Evaluations



- **NIST** (National Institute of Standards & Technology) conducts annual evaluation of speaker verification technology (since '95)
- **Aim:** Provide a common paradigm for comparing technologies
- **Focus:** Conversational telephone speech (text-independent)

Data Provider

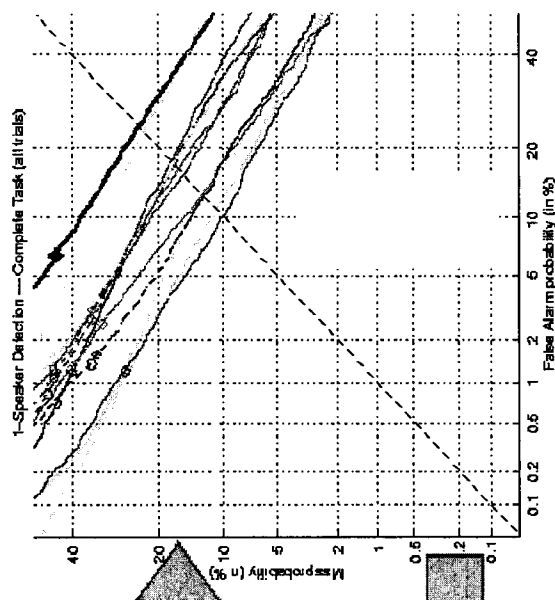


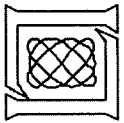
Evaluation Coordinator



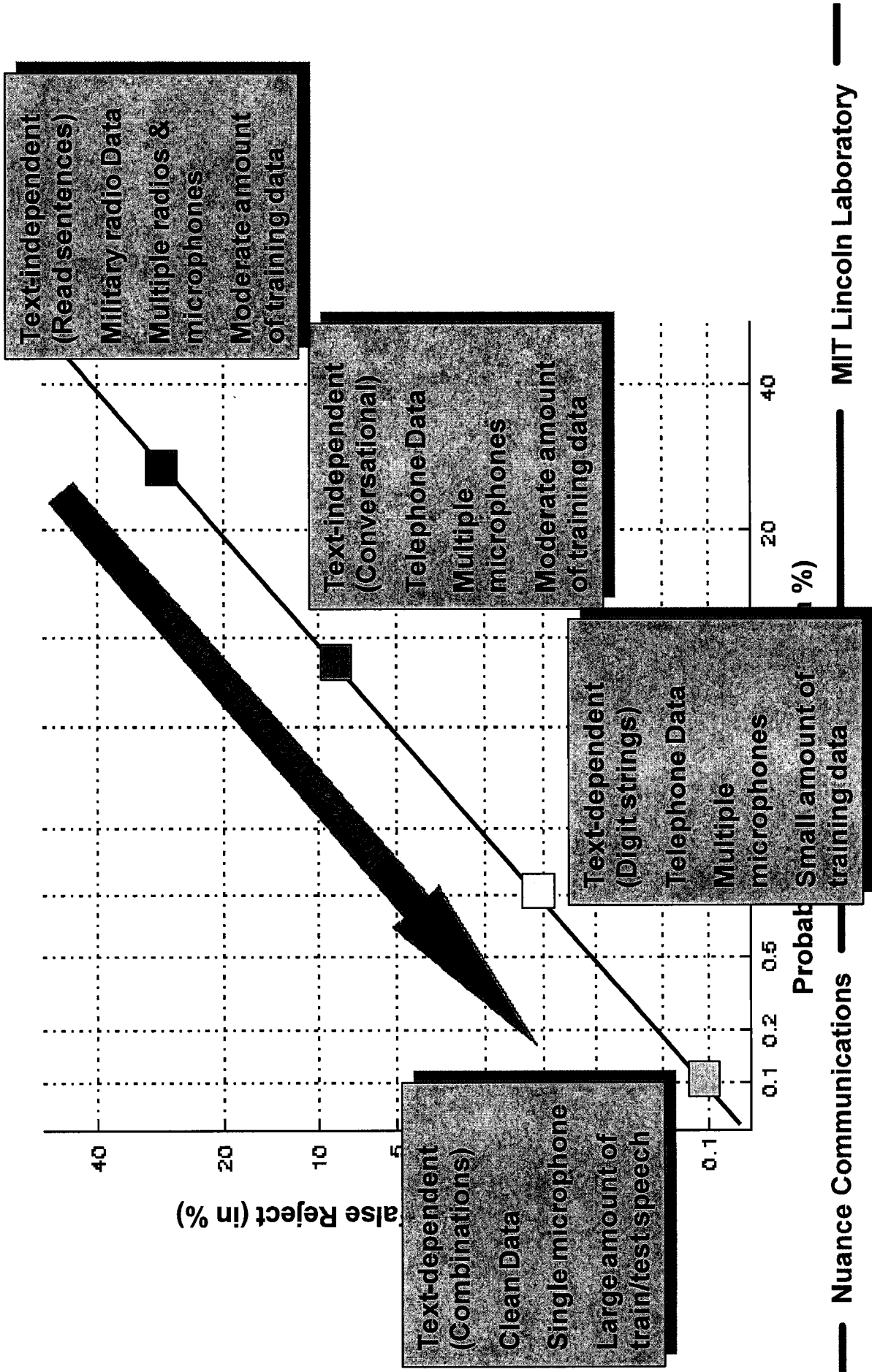
Comparison of
technologies on
common task

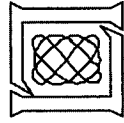
Technology Developers





Verification Performance Range of Performance





Verification Performance

Human vs. Machine

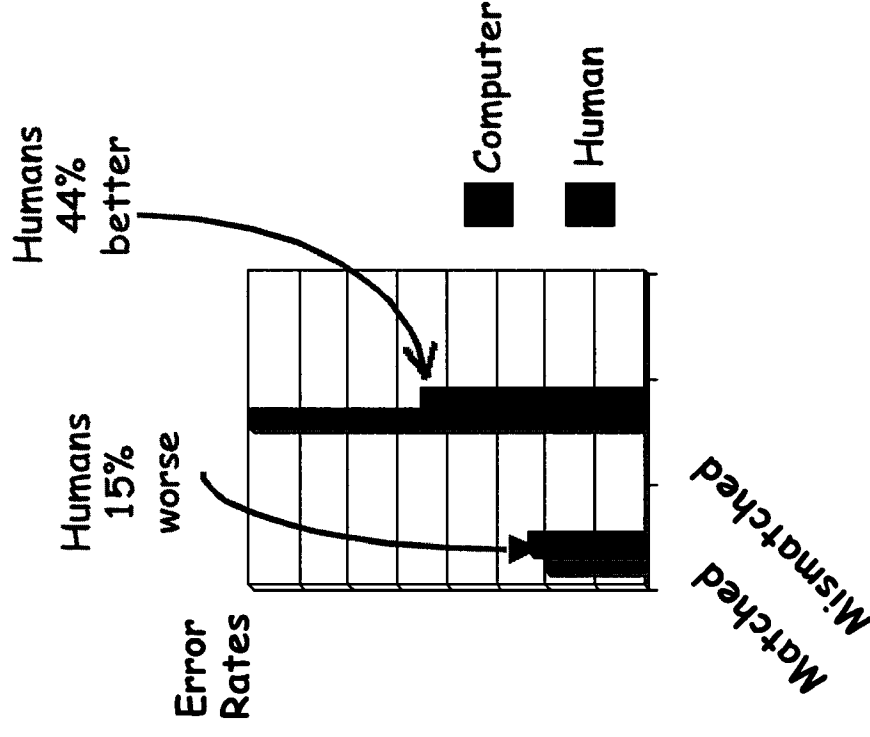


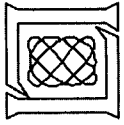
- **Motivation for comparing human to machine**

- Evaluating speech coders and potential forensic applications

- **Schmidt-Nielsen and Crystal used NIST evaluation (DSP Journal, January 2000)**

- Same amount of training data
- Matched Handset-type tests
- Mismatched Handset-type tests
- Used 3-sec conversational utterances from telephone speech

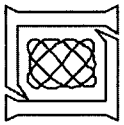




Outline



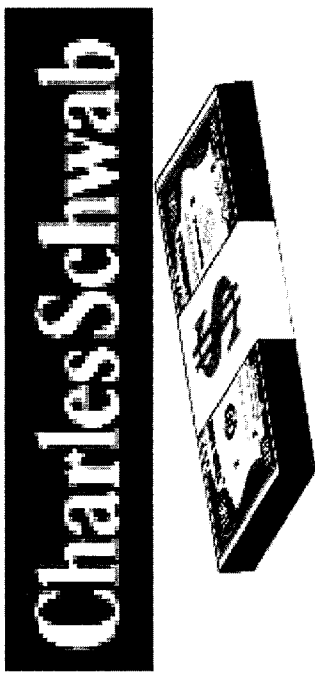
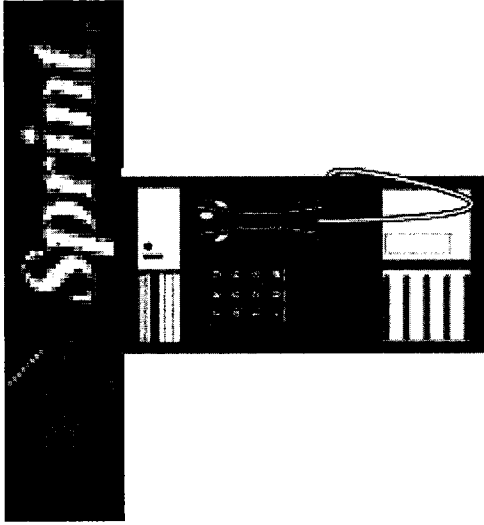
- Introduction
- General theory
- Performance
- Applications
- Conclusions and future directions

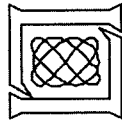


Applications



- Transaction authentication
 - Toll fraud prevention
 - Telephone credit card purchases
 - Telephone brokerage (e.g., stock trading)

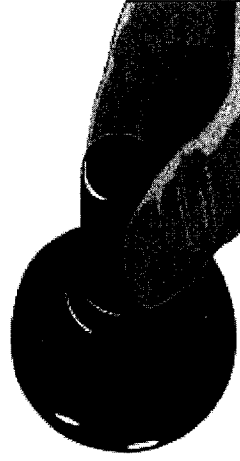




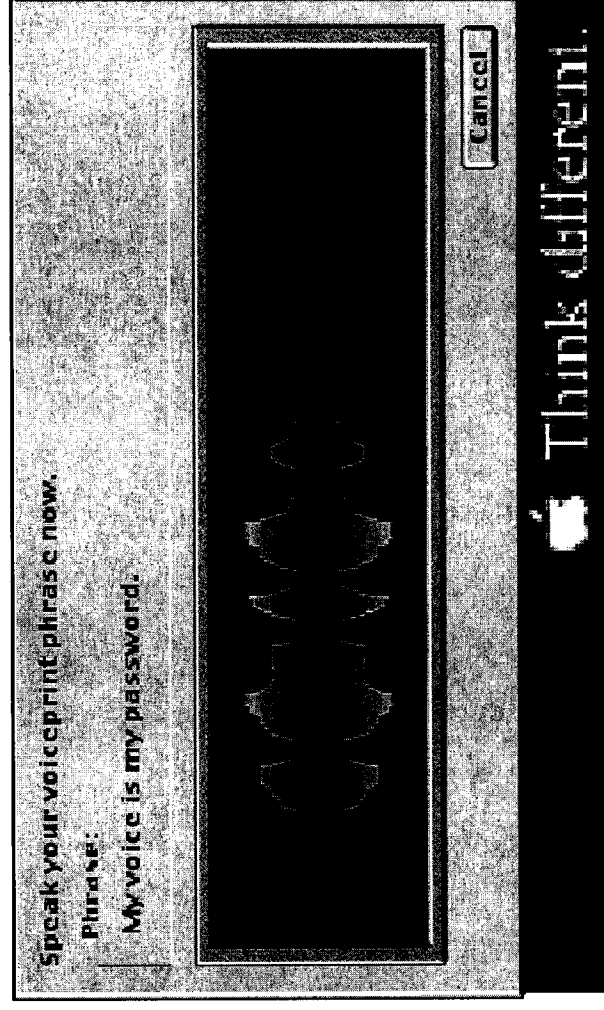
Applications

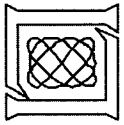


- **Access control**
 - Physical facilities
 - Computers and data networks



Mac OS9

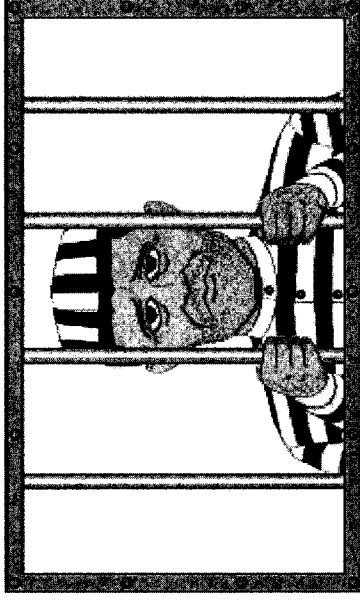


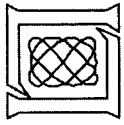


Applications



- **Monitoring**
 - Remote time and attendance logging
 - Home parole verification
 - Prison telephone usage

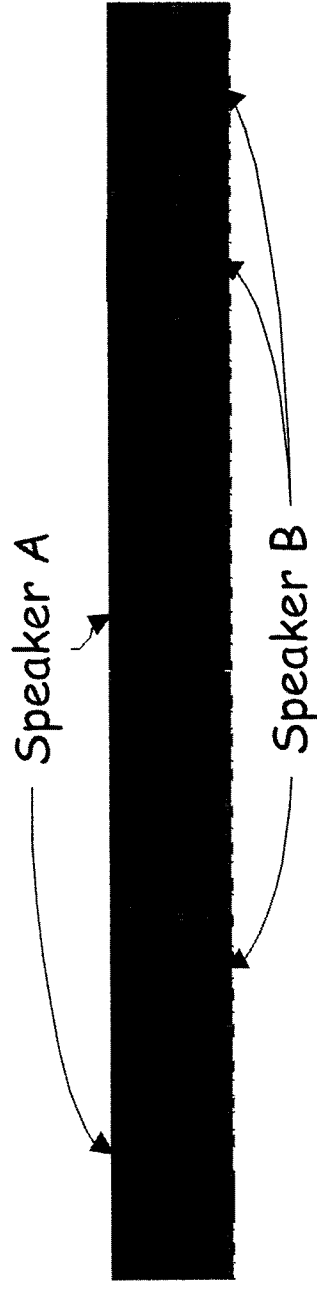
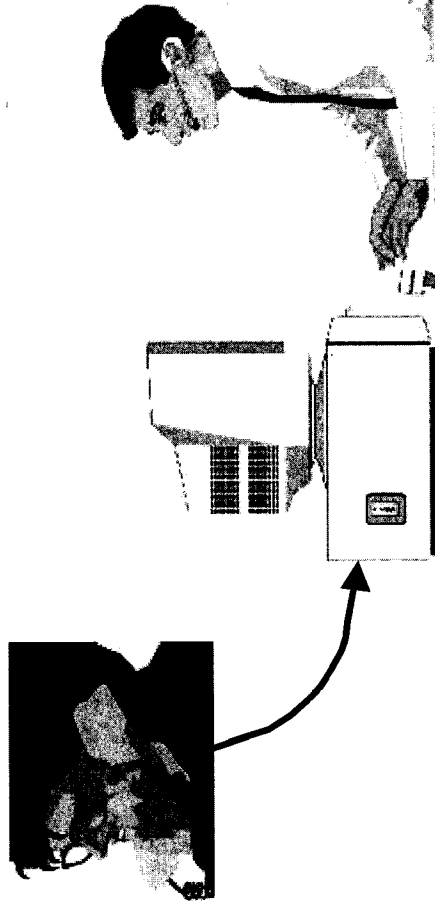


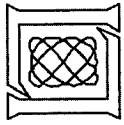


Applications



- **Information retrieval**
 - Customer information for call centers
 - Audio indexing (speech skimming device)

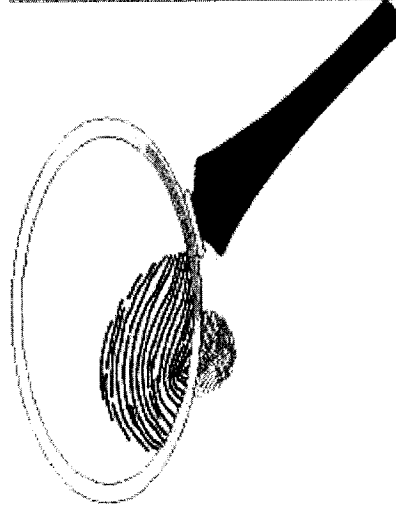




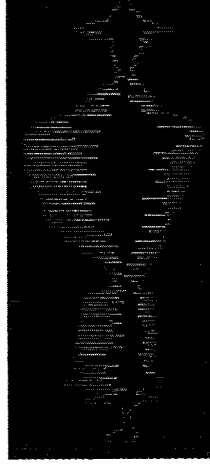
Applications



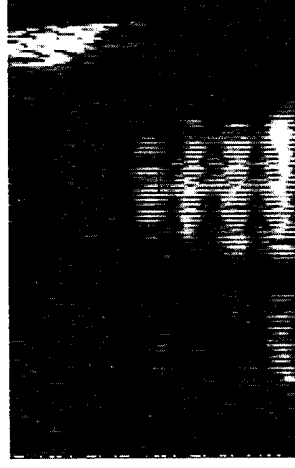
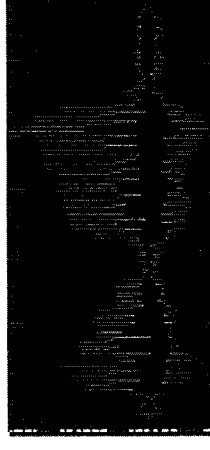
- **Forensics**
 - Voice sample matching

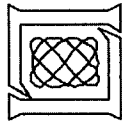


Recorded threat



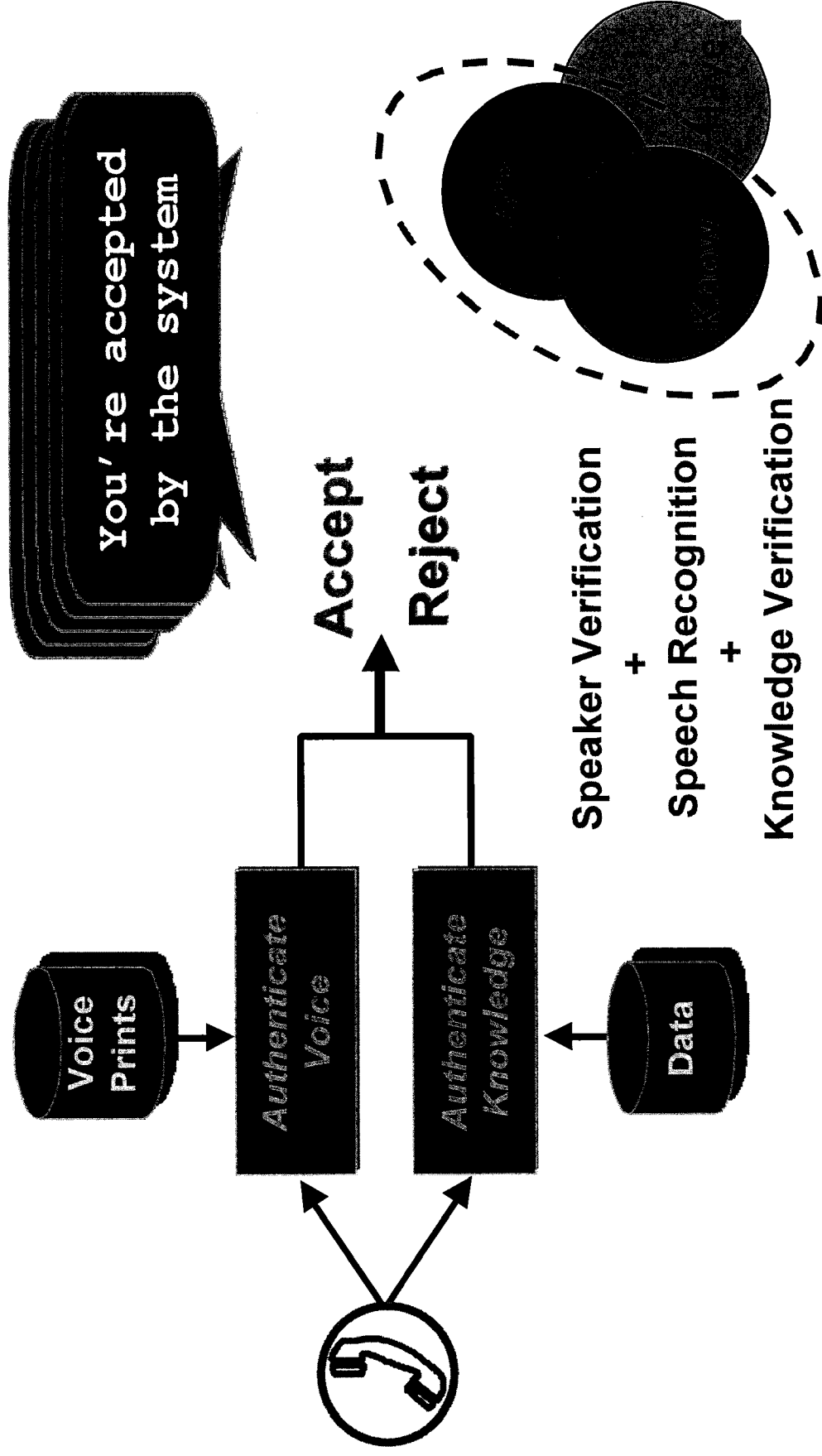
Suspect

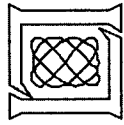




Applications

Speaker + Speech Recognition





Applications

First High-Volume Deployment



Benefits

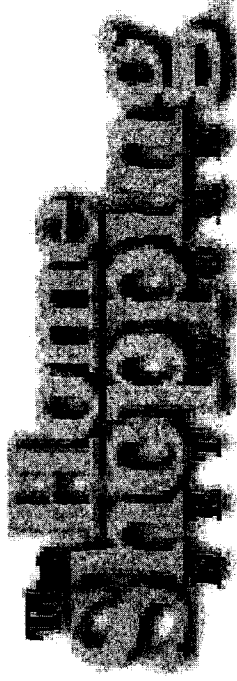
- Security
- Personalization

Application

- Speaker verification and identification based on home phone number

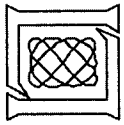
Size & Volume

- 250k customers enrolled currently @20K calls/day
- 5 million customers will enroll by Q2 '00 @170K calls/day



Implementation

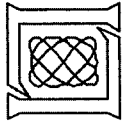
- Nuance Verifier™
- Edify telephony platform
- Deployed July 1999



Outline



- Introduction
- General theory
- Performance
- Applications
- Conclusions and future directions



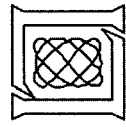
Conclusions



Speaker recognition is one of the few recognition areas where machines can outperform humans

Speaker recognition technology is a viable technique currently available for applications

Speaker recognition can be augmented with other authentication techniques to increase security



Future Directions



Research will focus on using speaker recognition for more unconstrained, uncontrolled situations

- Audio search and retrieval
- Increasing robustness to channel variability
- Incorporating higher-levels of knowledge into decisions

Speaker recognition technology will become an integral part of speech interfaces

- Personalization of services and devices
- Unobtrusive protection of transactions and information

Exhibit C



Welcome Kevin Kercher, IEEE Member

☐ Request Invalid
[BROWSE](#)[SEARCH](#)[IEEE XPLORE GUIDE](#)[SUPPORT](#)

The content you requested is not included in your subscription.

Login

Username

Password

» [Forgot your password?](#)

Please remember to log out when you have finished your session.

Access this document

- » [Buy this document now](#)
- » [Learn more about subscription options access information](#)
- » [Learn more about purchasing articles and standards](#)

Article Information**New techniques for automatic speaker verification**

Rosenberg, A.; Sambur, M.

Acoustics, Speech, and Signal Processing [see also IEEE Transactions on Signal Processing], IEEE Transactions on

Volume 23, Issue 2, Apr 1975 Page(s): 169 - 176

Digital Object Identifier

Summary: An interactive automatic speaker verification system has been augmented to include linear prediction parameters in addition to the already existing pitch and intensity analysis of sentence-long utterances. This improved system has been evaluated on a new and enlarged speaker population. A method for selecting optimum speaker-dependent features has been incorporated in this system which significantly improves its performance. The evaluation indicates that verification error rate is approximately 1 percent with respect to casual impostors and 4 percent with respect to well-trained mimics.

» [View citation and abstract](#)**IEEE Members**

Please review your [subscription information](#). If you have any questions, please complete the online [Technical Support Form](#).

If you are an IEEE Member Digital Library Subscriber and experiencing difficulty accessing your subscription, please check your [account profile](#).

IEEE Communications Society members: If you subscribe to the IEEE Electronic Periodicals Package or IEEE Electronic Periodicals Package Plus, you must access your subscription at www.comsoc.org.

Users at Subscribing Organizations

Please review your [subscription information](#). If you have any questions, please complete the online [Technical Support Form](#).

IEEE Computer Society Digital Library (CSLSP-e) subscribers: You must access your subscription at www.computer.org.

Already Purchased This Article?

Select the [Purchase History](#) link to access the document. You will have 5 days hours after purchase to access the Full Text PDF. Please complete the online [Technical Support Form](#) if you need assistance.

Guests

- Search and access Abstract records free of charge
- [Register](#) for table of contents alerts
- Purchase Full Text PDF documents

» [Learn more about subscription options](#) or how to become an [IEEE Member](#).**Message #RI2004**[Learn more about IEEE Subscriptions](#)

Indexed by
 Inspec®

[Help](#) [Contact Us](#) [Privacy & Security](#) [IEEE.org](#)

© Copyright 2008 IEEE – All Rights Reserved

Exhibit D



836,479,948 visitors served.

 TheFreeDictionary  Google


Automatic Speaker Verification









Search

?

 Word / Article  Starts with  Ends with  Text

subscription: ?

 Hutchinson encyclopedia

 Dictionary thesaurus	 Medical dictionary	 Legal dictionary	 Financial dictionary	 Acronyms	 Idioms	 Encyclopedia	 Wikipedia encyclopedia	?
--	--	--	--	--	--	--	--	---

ASV

(redirected from *Automatic Speaker Verification*)

0.06 sec.

Free Email Dictation

Deliver recorded phone dictation securely to your email.
www.messageshuttle.com

Ads by Google

Customer Intelligence

Utopy: Uncover the why behind Customer and Prospect behavior.
www.utopy.com

Keynote speaker

Motivate your audience with a professional keynote speaker.
www.DisneyInstitute.com

Page tools

 Printer friendly Cite / link

Clip Art

Dogs

Puppies

Chris Brown

Top Image Search

search better now

Source: Ask.com March 2008

Advertisement (you can turn it off)

Acronym Definition

ASV	Adaptive Supply Voltage
ASV	Advanced Safety Vehicle
ASV	Advanced Super View (Sharp LCD display)
ASV	Age Sexe Ville (French: Age, Sex, City)
ASV	Air Switching Valve
ASV	Air-To-Surface Vessel
ASV	Alkali Spreading Value
ASV	All Season Vehicle
ASV	Allgemeiner Sportverein (German Sports Association)
ASV	Amboseli, Kenya (airport code)
ASV	American Standard Version
ASV	Anodic Stripping Voltammetry
ASV	Application-Specific Vulnerability
ASV	Archivio Segreto Vaticano (Italian: Vatican Secret Archive)
ASV	Armored Security Vehicle
ASV	Autogenous Saphenous Vein
ASV	Automatic Self-Verification
ASV	Automatic Signature Verification
ASV	Automatic Speaker Verification
ASV	Average System Value
ASV	Avian Sarcoma Virus
ASV	Axel Springer Verlag (German publishing company)

[submit new definition](#)Powered by AcronymFinder.com © 1988-2007, Mountain Data Systems, All rights reserved.

Link to this

page: <http://acronyms.thefreedictionary.com/Automatic+Speaker+Verification>>ASV

Please bookmark with social media, your votes are noticed and appreciated:



Exhibit E

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Application of:	George Alfred Velius	Group No.:	2129
Serial No.:	09/886,824	Atty. Docket No.:	41942-52970
Filed:	June 21, 2001	Confirmation No.:	6850
		Customer No.:	021888
For:	Normalized Detector Scaling	Examiner:	Nathan H. Brown, Jr.

Commissioner of Patents and Trademarks
P.O. Box 1450
Alexandria, VA 22313-1450

DECLARATION OF DAVID P. MORGAN UNDER 37 C.F.R. §§ 131-132

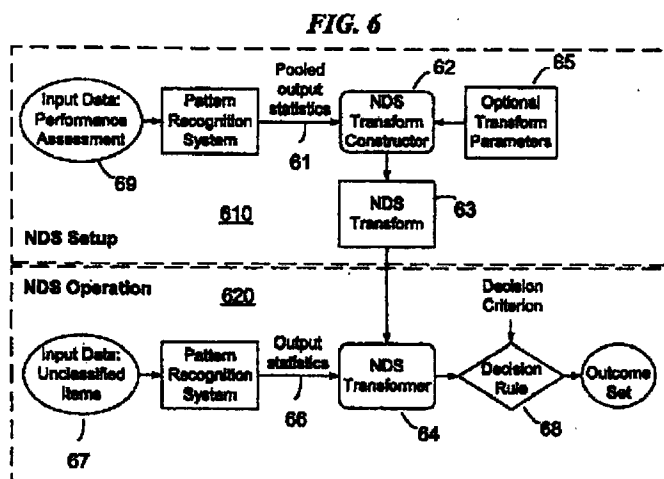
I, David P. Morgan, Ph.D., the below named Declarant, do hereby declare and state as follows:

1. My name is David P. Morgan, and I am the Vice President, Enterprise Technology & Architecture of Fidelity Investments Systems Company in Boston, Massachusetts.
2. Speaker identity verification (SIV) "engines" are sophisticated, computer-implemented systems used to enroll and subsequently verify a person's identity using the unique features of one's voice ("voice authentication" or "voice biometrics").
3. A person's voice, unlike other biometrics that measure static physical geometry such as fingerprints or iris scans, is affected by anatomical, physiological and behavioral factors. A person's voice also changes over time. As well, speech is affected by the voice interface utilized by the speaker (e.g., the microphone and electronics in a wire-line telephone or cellular phone) and the "network" effects of various telephone networks components. SIV engines, therefore, analyze a very large set of speech features and apply multi-dimensional statistical processing of submitted speech utterances to enroll and subsequently verify a speaker's identity.
4. A simple business example of the use of speaker identity verification in the financial services industry is the enrollment of an account holder's voice when opening an account with a financial institution (FI). Once enrolled, the account holder can be biometrically authenticated when calling the FI's self-service call center to check their account balance (once they have provided an "identity claim" such as their account number).

5. Most SIV engines provide an output in the form of a Yes/No decision about the authenticity of the person claiming to be the account holder, based upon the speech collected and submitted by the self-service call center application to the SIV engine. Alternatively, the SIV engine can return a “raw” numerical output score that can be used by the self-service application. These raw numbers can vary widely for a given speaker from one utterance to another, and from one call to another call – which can lead to a higher number of instances in which the call center application falsely accepts an impostor calling as the account holder, as well as a higher number of false rejections of the authentic account holder. False accepts (FA’s) and false rejects (FR’s) both have significant business implications for the FI.
6. In the financial services industry, the FI has significant regulatory, financial and reputation risks associated with an impostor accessing an authentic customer’s account. For example, the U.S. Congress’s Gramm-Leach-Bliley Act requires FI’s to strongly protect the privacy of personally-identifiable financial information. The FDIC and FFIEC have published guidelines that require FI’s to assess the risk of impostors gaining access to their customer’s financial information over the telephone or Internet, and implement appropriate measures to mitigate those risks (e.g., “multi-factor authentication” to gain access to account information).
7. At the same time, an FI has to ensure that the services provided to its customers are as convenient as possible. Convenient customer service is essential for customer acquisition and retention in a very competitive marketplace. For that reason, self-service channels such as automated telephone systems, mobile applications and web-based services are being rapidly implemented.
8. Therefore, it is essential that FI’s implement customer service methods that balance the security/risk requirements with customer convenience. In these types of self-service transactions, FI’s can actually weigh the cost of false accept (e.g., a literal financial loss due to an impostor) with the cost of a false reject (e.g., customer inconvenience, dissatisfaction, or loss of the customer to another FI). FI’s can consider a variety of risk factors for a transaction (e.g., risk for each different type of transaction; dollar value of the transaction; history with that particular customer; etc.) to assist in adjusting the security-convenience tradeoff for that particular transaction. In the example of an FI account holder, simple access to their account information would typically be a low risk transaction that would not require a high level of security. An account holder attempting to make a \$5,000 wire transfer would require a higher level of security; a \$50,000 wire

transfer would require even higher security and the customer would likely understand and appreciate a higher level of security in these transactions.

9. FI's use a variety of risk assessment methods, and many use scoring systems which provide decision support for a particular transaction (e.g., credit scores for loans; "velocity" scores for credit card transactions). Those scores are used in conjunction with other pertinent information and business rules established by the FI to decide whether to accept the transaction or perform other actions.
10. The following is an example of implementing the Velius Normalized Detector Scaling invention (NDS), as described in U.S. Patent Application Publication No. 2002/0198857, to provide a breakthrough advance in financial services applications of SIV. The SIV engines used are non-parametric pattern recognition systems as described in Velius and shown in Figure 6 below.

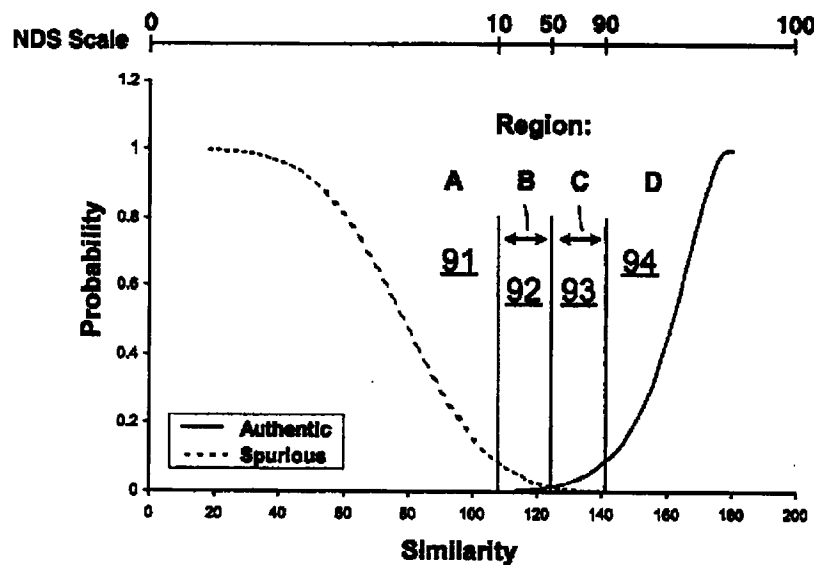


11. As illustrated in the NDS Setup in Velius Fig. 6, the **Pooled output statistics** (raw SIV verification scores that are known authentic and known spurious) from the SIV engine (61) and the optional **Transform Parameters** (65) [examples described below] can be used by a computer-implemented method [the **NDS Transform Constructor** (62)] to create the **NDS Transform** (63).
12. **Example Transform Parameter 1:** The scores range from 1 to 99, where 1 represents almost no chance that the speaker is the authentic account holder, and 99 represents almost no chance that the speaker is not the authentic account holder.
13. **Example Transform Parameter 2:** The mid-point of the scale (a score of 50) is calibrated to represent the "Equal Error Rate" (EER), where there is an equal chance that the

speaker is not the authentic account holder and an equal chance that the speaker is the authentic account holder. This “normalization” of the confidence score scale is critical to implementing business rules that don’t change over time as the underlying statistics from the SIV engine may change.

14. Example Transform Parameter 3: The majority of the resolution of the scale (10 to 90) covers the range in which there is the greatest chance of a false acceptance of an impostor or a false rejection of the authentic account holder.
15. The following is an example of Velius Fig. 9 that illustrates the implementation of the three Transform Parameters specified above.

FIG. 9



16. In the **NDS Operation** (see Fig. 6 above), unclassified, raw verification scores from the output of the SIV engine (66) are transformed in real time by a computer-implemented method [the **NDS Transformer** (64)] onto the one-dimensional, NDS “confidence score” scale (1 – 99) illustrated in Fig. 9 above (note the example **NDS Scale** superimposed at the top of Fig. 9).
17. Both the NDS Set-up and the NDS Operation processes require a computer; a person skilled in this area would know that manual calculations would be impossible.
18. Financial services companies are implementing telephone-based applications that use a voice authentication service that verifies the identity of an account holder using the Normalized Detector Scaling method described above. The ability for companies to implement sophisticated business rules that “tune” the security-convenience tradeoff in real-time on a transaction-by-transaction basis using the risk characteristics specific to

each transaction is both extremely powerful and critically important for mitigating financial and regulatory compliance risks. I expect that the Normalized Detector Scaling will become commonplace in implementing SIV in financial services. A 2007 report published by Opus Research has estimated that voice authentication market revenues will exceed \$700 million by 2011.

19. Normalized Detector Scaling is a significant advance in the field of complex, non-parametric pattern recognitions systems, such as speaker identity verification systems, in that it enables decision rules to be established in a way that is independent of the features, or the particular statistics, employed by the pattern recognition system.
20. The term "adaptive speaker identity verification system," found on Page 2, Paragraph [0016], Lines 1-2 of the original patent application of George Alfred Velius, i.e., U.S. Patent Application Publication No. 2002/0198857, published December 26, 2002, is well known in the art for a physical machine, having a computer, that receives a person's unclassified speech and converts that speech to data and then is able to perform analysis on that data utilizing statistics to verify the identity of a particular person. A person skilled in speaker identity verification technology would be easily able to implement the Applicant's Invention disclosed in U.S. Patent Application Publication No. 2002/0198857 in an adaptive speaker identity verification system by merely reading U.S. Patent Application Publication No. 2002/0198857 and then programming the adaptive speaker identity verification system. This is a very straight forward process based on my reading of U.S. Patent Application Publication No. 2002/0198857 so there would be no need for any undue experimentation involving the adaptive speaker identity verification system.
21. I further declare that all statements made herein by my own knowledge are true and all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the above-identified application.

Further Declarant Sayeth Not.

March 31, 2008

Date



Name: David P. Morgan

Exhibit F

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Application of:	George Alfred Velius	Group No.:	2129
Serial No.:	09/886,824	Atty. Docket No.:	41942-52970
Filed:	June 21, 2001	Confirmation No.:	6850
		Customer No.:	021888
For:	Normalized Detector Scaling	Examiner:	Nathan H. Brown, Jr.

Commissioner of Patents and Trademarks
P.O. Box 1450
Alexandria, VA 22313-1450

DECLARATION OF MICHAEL PHILLIPS UNDER 37 C.F.R. §§ 131-132

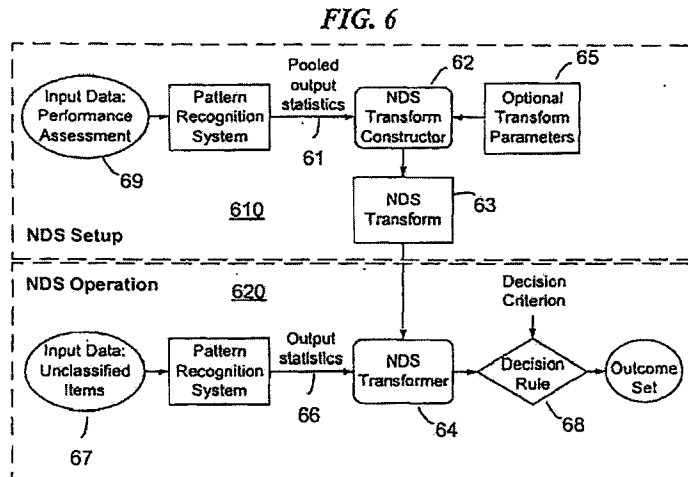
I, Michael Phillips, the below named Declarant, do hereby declare and state as follows:

1. My name is Michael Phillips, and I am the Co-Founder and Chief Technology Officer of vlingo, in Cambridge, Massachusetts, and the Co-Founder of SpeechWorks International in 1994 (now Nuance Communications, Inc., Boston, Massachusetts). I currently serve as a member of the TradeHarbor Advisory Board.
2. Speaker identity verification (SIV) "engines" are sophisticated, computer-implemented systems used to enroll and subsequently verify a person's identity using the unique features of one's voice ("voice authentication" or "voice biometrics").
3. A person's voice, unlike other biometrics that measure static physical geometry such as fingerprints or iris scans, is affected by anatomical, physiological and behavioral factors. A person's voice also changes over time. As well, speech is affected by the voice interface utilized by the speaker (e.g., the microphone and electronics in a wire-line telephone or cellular phone) and the "network" effects of various telephone networks components. SIV engines, therefore, analyze a very large set of speech features and apply multi-dimensional statistical processing of submitted speech utterances to enroll and subsequently verify a speaker's identity.
4. A simple business example of the use of speaker identity verification in the financial services industry is the enrollment of an account holder's voice when opening an account with a financial institution (FI). Once enrolled, the account holder can be biometrically authenticated when calling the FI's self-service call center to check their account balance (once they have provided an "identity claim" such as their account number).

5. Most SIV engines provide an output in the form of a Yes/No decision about the authenticity of the person claiming to be the account holder, based upon the speech collected and submitted by the self-service call center application to the SIV engine. Alternatively, the SIV engine can return a “raw” numerical output score that can be used by the self-service application. These raw numbers can vary widely for a given speaker from one utterance to another, and from one call to another call – which can lead to a higher number of instances in which the call center application falsely accepts an impostor calling as the account holder, as well as a higher number of false rejections of the authentic account holder. False accepts (FA’s) and false rejects (FR’s) both have significant business implications for the FI.
6. In the financial services industry, the FI has significant regulatory, financial and reputation risks associated with an impostor accessing an authentic customer’s account. For example, the U.S. Congress’s Gramm-Leach-Bliley Act requires FI’s to strongly protect the privacy of personally-identifiable financial information. The FDIC and FFIEC have published guidelines that require FI’s to assess the risk of impostors gaining access to their customer’s financial information over the telephone or Internet, and implement appropriate measures to mitigate those risks (e.g., “multi-factor authentication” to gain access to account information).
7. At the same time, an FI has to ensure that the services provided to its customers are as convenient as possible. Convenient customer service is essential for customer acquisition and retention in a very competitive marketplace. For that reason, self-service channels such as automated telephone systems, mobile applications and web-based services are being rapidly implemented.
8. Therefore, it is essential that FI’s implement customer service methods that balance the security/risk requirements with customer convenience. In these types of self-service transactions, FI’s can actually weigh the cost of false accept (e.g., a literal financial loss due to an impostor) with the cost of a false reject (e.g., customer inconvenience, dissatisfaction, or loss of the customer to another FI). FI’s can consider a variety of risk factors for a transaction (e.g., risk for each different type of transaction; dollar value of the transaction; history with that particular customer; etc.) to assist in adjusting the security-convenience tradeoff for that particular transaction. In the example of an FI account holder, simple access to their account information would typically be a low risk transaction that would not require a high level of security. An account holder attempting to make a \$5,000 wire transfer would require a higher level of security; a \$50,000 wire

transfer would require even higher security and the customer would likely understand and appreciate a higher level of security in these transactions.

9. FI's use a variety of risk assessment methods, and many use scoring systems which provide decision support for a particular transaction (e.g., credit scores for loans; "velocity" scores for credit card transactions). Those scores are used in conjunction with other pertinent information and business rules established by the FI to decide whether to accept the transaction or perform other actions.
10. The following is an example of implementing the Velius Normalized Detector Scaling invention (NDS), as described in U.S. Patent Application Publication No. 2002/0198857, to provide a breakthrough advance in financial services applications of SIV. The SIV engines used are non-parametric pattern recognition systems as described in Velius and shown in Figure 6 below.

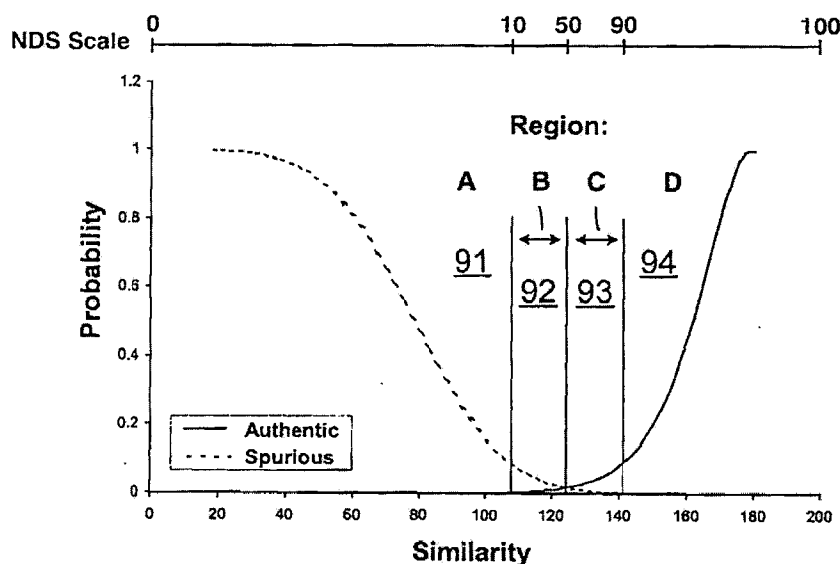


11. As illustrated in the **NDS Setup** in Velius Fig. 6, the **Pooled output statistics** (raw SIV verification scores that are known authentic and known spurious) from the SIV engine (61) and the optional **Transform Parameters** (65) [examples described below] can be used by a computer-implemented method [the **NDS Transform Constructor** (62)] to create the **NDS Transform** (63).
12. Example Transform Parameter 1: The scores range from 1 to 99, where 1 represents almost no chance that the speaker is the authentic account holder, and 99 represents almost no chance that the speaker is not the authentic account holder.
13. Example Transform Parameter 2: The mid-point of the scale (a score of 50) is calibrated to represent the "Equal Error Rate" (EER), where there is an equal chance that the

speaker is not the authentic account holder and an equal chance that the speaker is the authentic account holder. This “normalization” of the confidence score scale is critical to implementing business rules that don’t change over time as the underlying statistics from the SIV engine may change.

14. Example Transform Parameter 3: The majority of the resolution of the scale (10 to 90) covers the range in which there is the greatest chance of a false acceptance of an impostor or a false rejection of the authentic account holder.
15. The following is an example of Velius Fig. 9 that illustrates the implementation of the three Transform Parameters specified above.

FIG. 9



16. In the **NDS Operation** (see Fig. 6 above), unclassified, raw verification scores from the output of the SIV engine (66) are transformed in real time by a computer-implemented method [the **NDS Transformer** (64)] onto the one-dimensional, NDS “confidence score” scale (1 – 99) illustrated in Fig. 9 above (note the example **NDS Scale** superimposed at the top of Fig. 9).
17. Both the NDS Set-up and the NDS Operation processes require a computer; a person skilled in this area would know that manual calculations would be impossible.
18. Financial services companies are implementing telephone-based applications that use a voice authentication service that verifies the identity of an account holder using the Normalized Detector Scaling method described above. The ability for companies to implement sophisticated business rules that “tune” the security-convenience tradeoff in real-time on a transaction-by-transaction basis using the risk characteristics specific to

each transaction is both extremely powerful and critically important for mitigating financial and regulatory compliance risks. I expect that the Normalized Detector Scaling will become commonplace in implementing SIV in financial services. A 2007 report published by Opus Research has estimated that voice authentication market revenues will exceed \$700 million by 2011.

19. Normalized Detector Scaling is a significant advance in the field of complex, non-parametric pattern recognitions systems, such as speaker identity verification systems, in that it enables decision rules to be established in a way that is independent of the features, or the particular statistics, employed by the pattern recognition system.
20. The term "adaptive speaker identity verification system," found on Page 2, Paragraph [0016], Lines 1-2 of the original patent application of George Alfred Velius, i.e., U.S. Patent Application Publication No. 2002/0198857, published December 26, 2002, is well known in the art for a physical machine, having a computer, that receives a person's unclassified speech and converts that speech to data and then is able to perform analysis on that data utilizing statistics to verify the identity of a particular person. A person skilled in speaker identity verification technology would be easily able to implement the Applicant's Invention disclosed in U.S. Patent Application Publication No. 2002/0198857 in an adaptive speaker identity verification system by merely reading U.S. Patent Application Publication No. 2002/0198857 and then programming the adaptive speaker identity verification system. This is a very straight forward process based on my reading of U.S. Patent Application Publication No. 2002/0198857 so there would be no need for any undue experimentation involving the adaptive speaker identity verification system.
21. I further declare that all statements made herein by my own knowledge are true and all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the above-identified application.

Further Declarant Sayeth Not.

March 31 2008

Date



Name: Michael Phillips